

ENTERICALLY TRANSMITTED NON-A/NON-B HEPATITIS  
VIRAL AGENT AND CHARACTERISTIC EPITOPES THEREOF

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation of U.S. Application Serial No. 08/279,823, filed July 25, 1994, which is a continuation of U.S. Application Serial No. 07/681,078, filed April 5, 1991, now abandoned, which is a continuation-in-part of U.S. Application Serial No. 07/505,888, filed April 5, 1990, now abandoned, which is a continuation-in-part of U.S. Application Serial No. 07/420,921, filed October 13, 1989, now abandoned, which is a continuation-in-part of U.S. Application Serial No. 07/367,486, filed June 16, 1989, now abandoned, which is a continuation-in-part of U.S. Application Serial No. 07/336,672, filed April 11, 1989, now abandoned, which is a continuation-in-part of U.S. Application Serial No. 07/208,997, filed June 17, 1988, now abandoned, all of which are herein incorporated by reference.

INTRODUCTION

Field of Invention

This invention relates to recombinant proteins, genes, and gene probes and more specifically to such proteins and probes derived from an enterically transmitted nonA/nonB hepatitis viral agent, to diagnostic methods and vaccine applications which employ the proteins and probes, and to gene segments that encode specific epitopes (and proteins artificially produced to contain those epitopes) that are particularly useful in diagnosis and prophylaxis.

Background

Enterically transmitted non-A/non-B hepatitis viral agent (ET-NANB; also referred to herein as HEV) is the reported cause of hepatitis in several epidemics and sporadic cases in Asia, Africa, Europe, Mexico, and the Indian subcontinent. Infection is usually by water contaminated with feces, although

the virus may also spread by close physical contact.  
The virus does not seem to cause chronic infection.  
The viral etiology in ET-NANB has been demonstrated by  
infection of volunteers with pooled fecal isolates;  
5 immune electron microscopy (IEM) studies have shown  
virus particles with 27-34 nm diameters in stools  
from infected individuals. The virus particles reacted  
with antibodies in serum from infected individuals  
from geographically distinct regions, suggesting that  
10 a single viral agent or class is responsible for the  
majority of ET-NANB hepatitis seen worldwide. No  
antibody reaction was seen in serum from individuals  
infected with parenterally transmitted NANB virus  
(also known as hepatitis C virus or HCV), indicating  
15 a different specificity between the two NANB types.

In addition to serological differences, the  
two types of NANB infection show distinct clinical  
differences. ET-NANB is characteristically an acute  
infection, often associated with fever and arthralgia,  
20 and with portal inflammation and associated bile  
stasis in liver biopsy specimens (Arankalle).  
Symptoms are usually resolved within six weeks.  
Parenterally transmitted NANB, by contrast, produces a  
chronic infection in about 50% of the cases. Fever and  
arthralgia are rarely seen, and inflammation has a  
25 predominantly parenchymal distribution (Khuroo, 1980).  
The course of ET-NANBH is generally uneventful in  
healthy individuals, and the vast majority of those  
infected recover without the chronic sequelae seen  
with HCV. One peculiar epidemiologic feature of this  
30 disease, however, is the markedly high mortality  
observed in pregnant women; this is reported in  
numerous studies to be on the order of 10-20%. This  
finding has been seen in a number of epidemiologic  
35 studies but at present remains unexplained. Whether  
this reflects viral pathogenicity, the lethal  
consequence of the interaction of virus and immune  
suppressed (pregnant) host, or a reflection of the

debilitated prenatal health of a susceptible  
malnourished population remains to be clarified.

The two viral agents can also be distinguished on the basis of primate host susceptibility.  
5 ET-NANB, but not the parenterally transmitted agent,  
can be transmitted to cynomolgus monkeys. The  
parenterally transmitted agent is more readily  
transmitted to chimpanzees than is ET-NANB (Bradley,  
1987).

10 There have been major efforts worldwide to  
identify and clone viral genomic sequences associated  
with ET-NANB hepatitis. One goal of this effort,  
requiring virus-specific genomic sequences, is to  
identify and characterize the nature of the virus and  
15 its protein products. Another goal is to produce  
recombinant viral proteins which can be used in  
antibody-based diagnostic procedures and for a  
vaccine. Despite these efforts, viral sequences  
associated with ET-NANB hepatitis have not been  
20 successfully identified or cloned heretofore, nor have  
any virus-specific proteins been identified or  
produced.

#### Relevant Literature

- 25 Arankalle, V.A., et al., The Lancet, 550  
(March 12, 1988).  
Bradley, D.W., et al., J Gen. Virol., 69:1  
(1988).  
Bradley, D.W. et al., Proc. Nat. Acad. Sci.,  
30 USA, 84:6277 (1987).  
Gravelle, C.R. et al., J. Infect. Diseases,  
131:167 (1975).  
Kane, M.A., et al., JAMA, 252:3140 (1984).  
Khuroo, M.S., Am. J. Med., 48:818 (1980).  
35 Khuroo, M.S., et al., Am. J. Med., 68:818  
(1983).

Maniatis, T., et al. Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory (1982).

Seto, B., et al., *Lancet*, 11:941 (1984).

5 Sreenivasan, M.A., et al., *J. Gen. Virol.*, 65:1005 (1984).

Tabor, E., et al., *J. Infect. Dis.*, 140:789 (1979).

10 SUMMARY OF THE INVENTION

Novel compositions, as well as methods of preparation and use of the compositions are provided, where the compositions comprise viral proteins and fragments thereof derived from the viral agent for ET-  
15 NANB. A number of specific fragments of viral proteins (and the corresponding genetic sequences) that are particularly useful in diagnosis and vaccine production are also disclosed. Methods for preparation of ET-NANB viral proteins include isolating ET-NANB  
20 genomic sequences which are then cloned and expressed in a host cell. The resultant recombinant viral proteins find use as diagnostic agents and as vaccines. The genomic sequences and fragments thereof find use in preparing ET-NANB viral proteins and as  
25 probes for virus detection.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows vector constructions and manipulations used in obtaining and sequencing cloned  
30 ET-NANB fragment; and

Figures 2A-2B are representations of Southern blots in which a radiolabeled ET-NANB probe was hybridized with amplified cDNA fragments prepared from RNA isolated from infected (I) and non-infected  
35 (N) bile sources (2A), and from infected (I) and non-infected (N) stool-sample sources (2B).

## DESCRIPTION OF SPECIFIC EMBODIMENTS

Novel compositions comprising generic sequences and fragments thereof derived from the viral agent for ET-NANB are provided, together with  
5 recombinant viral proteins produced using the genomic sequences and methods of using these compositions. Epitopes on the viral protein have been identified that are particularly useful in diagnosis and vaccine production. Small peptides containing the epitopes are  
10 recognized by multiple sera of patients infected with ET-NANB.

The molecular cloning of HEV was accomplished by two very different approaches. The first successful identification of a molecular clone was  
15 based on the differential hybridization of putative HEV cDNA clones to heterogeneous cDNA from infected and uninfected cyno bile. cDNAs from both sources were labeled to high specific activity with  $^{32}\text{P}$  to identify a clone that hybridized specifically to the  
20 infected source probe. A cyno monkey infected with the Burma isolate of HEV was used in these first experiments. The sensitivity of this procedure is directly related to the relative abundance of the specific sequence against the overall background. In  
25 control experiments, it was found that specific identification of a target sequence may be obtained with as little as 1 specific part per 1000 background sequences. A number of clones were identified by this procedure using libraries and probes made from  
30 infected (Burma isolate) and control uninfected cyno bile. The first extensively characterized clone of the 16 plaques purified by this protocol was given the designation ET1.1.

ET1.1 was first characterized as both  
35 derived from and unique to the infected source cDNA. Heterogeneous cDNA was amplified from both infected and uninfected sources using a sequence independent single premier amplification technique (SISPA). This

technique is described in copending application serial  
No. 208,512, filed June 17, 1988. The limited pool of  
cDNA made from Burma infected cyno bile could then be  
amplified enzymatically prior to cloning or  
5 hybridization using putative HEV clones as probes.  
ET1.1 hybridized specifically to the original bile  
cDNA from the infected source. Further validation of  
this clone as derived from the genome of HEV was  
demonstrated by the similarity of the ET1.1 sequence  
10 and those present in SISPA cDNA prepared from five  
different human stool samples collected from  
different ET-NANBH epidemics including Somalia,  
Tashkent, Borneo, Mexico and Pakistan. These  
molecular epidemiologic studies established the  
15 isolated sequence as derived from the virus that  
represented the major cause of ET-NANBH worldwide.

The viral specificity of ET1.1 was further  
established by the finding that the clone hybridized  
specifically to RNA extracted from infected cyno  
20 liver. Hybridization analysis of polyadenylated RNA  
demonstrated a unique 7.5 Kb polyadenylated  
transcript not present in uninfected liver. The size  
of this transcript suggested that it represented the  
full length viral genome. Strand specific  
25 oligonucleotides were also used to probe viral genomic  
RNA extracted directly from semi-purified virions  
prepared from human stool. The strand specificity was  
based on the RNA-directed RNA polymerase (RDRP) open  
reading frame (ORF) identified in ET1.1 (see below).  
30 Only the probe detecting the sense strand hybridized  
to the nucleic acid. These studies characterized HEV  
as a plus sense, single stranded genome. Strand  
specific hybridization to RNA extracted from the liver  
also established that the vast majority of  
35 intracellular transcript was positive sense. Barring  
any novel mechanism for virus expression, the negative  
strand, although not detectable, would be present at a

ratio of less than 1:100 when compared with the sense strand.

ET1.1 was documented as exogenous when tested by both Southern blot hybridization and PCR using genomic DNAs derived from uninfected humans, infected and uninfected cynos and also the genomic DNAs from E. coli and various bacteriophage sources. The latter were tested in order to rule out trivial contamination with an exogenous sequence introduced during the numerous enzymatic manipulations performed during cDNA construction and amplification. It was also found that the nucleotide sequence of the ET1.1 clone was not homologous to any entries in the Genebank database. The translated open reading frame of the ET1.1 clone did, however, demonstrate limited homology with consensus amino acid residues consistent with an RNA-directed RNA polymerase. This consensus amino acid motif is shared among all positive strand RNA viruses and, as noted above, is present at the 3' end of the HCV genome. The 1.3 Kb clone was therefore presumed to be derived, at least in part, from the nonstructural portion of the viral genome.

Because of the relationship of different strains of ET-NANB to each other that has been demonstrated by the present invention, the genome of the ET-NANB viral agent is defined in this specification as containing a region which is homologous to the 1.33 kb DNA EcoRI insert present in plasmid pTZKF1 (ET1.1) carried in E. coli strain BB4 and having ATCC deposit no. 67717. The entire sequence, in both directions, has now been identified as set forth below. The sequences of both strands are provided, since both strands can encode proteins. However, the sequence in one direction has been designated as the "forward" sequence because of statistical similarities to known proteins and because the forward sequence is known to be predominately protein-encoding. This sequence is set forth below

along with the three possible translation sequences. There is one long open reading frame that starts at nucleotide 145 with an isoleucine and extends to the end of the sequence. The two other reading frames have many termination codons. Standard abbreviations for nucleotides and amino acids are used here and elsewhere in this specification.

The gene sequence given below is substantially identical to one given in the parent application. The present sequence differs in the omission of the first 37 nucleotides at the 5' end and last 13 nucleotides at the 3' end, which are derived from the linker used for cloning rather than from the virus. In addition, a G was omitted at position 227 of the sequence given in the parent application.

The following gene sequence has SEQ ID NO.1; the first amino acid sequence in reading frame beginning with nucleotide 1 has SEQ ID NO.2; the second amino acid sequence in reading frame beginning with nucleotide 2 has SEQ ID NO.3; and the third amino acid sequence in reading frame beginning with nucleotide 3 has SEQ ID NO.4.

#### Forward Sequence

##### SEQ ID NO. 1:

25	AGACCTGTCC CTGTTGCAGC TGTTCTACCA CCCTGCCCGG AGCTCGAACA GGGCCTTCTC	60
	TACCTGCCCC AGGAGCTCAC CACCTGTGAT AGTGTGTAA CATTGAATT AACAGACATT	120
30	GTGCACTGCC GCATGGCCGC CCGAGGCCAG CGCAAGGCCG TGCTGTCCAC ACTCGTGGGC	180
	CGCTACGGCG GTGCGACAAA GCTGTACAA GTTCCCACT CTGATGTTGG CGACTCTCTC	240
	GCCCGTTTTA TCCCGGCCAT TGCGCCGCTA CAGGTACAA CTGTGAATT GTACGAGCTA	300
35	GTGGAGGCCA TGGTCGAGAA GGGCGAGAT GGTCCGCCG TCCTTGAGCT TGATCTTTGC	360
	AACCGTGACG TGTCAGGAT CACCTCTTC CAGAAAGATT GTAACAAGTT CACCACAGGT	420
40	GAGACCATTG CCCATGGTAA AGTGGGCTAG GGCATCTCGG CCGGAGCAA GACCTTCTGC	480
	GCCCTCTTTG GCCCTTGGTT CCGGCTATT CAGAAAGCTA TTCTGGCCCT GCTCCCTCAG	540
45	GGTGTGTTTT ACGGTGATGC CTTTGATGAG ACCGTCTCT CCGCGGCTGT GGCCGCAGCA	600

	GCATCCA TGGTGTGGA GAATGACTTT TCTGAGTTTG	660
	TCTCTGGGTC TAGAGTGTG TATTATGGAG GAGTGTGGGA TGCCGCAGTG GCTCATCCGC	720
5	CTGTATCACC TTATAAGGTC TGGGTGGATC TTGCAGGCCC CGAAGGAGTC TCTGCGAGGG	780
	TTTTGGAAGA AACACTCCGG TGAGCCCGGC ACTTTCTAT GGAATACTGT CTGGAATATG	840
	GCCGTTATTA CCACTGTGA TGACTTCGGC GATTTTCAGG TGGCTGCCTT TAAAGGTGAT	900
10	GATTCGATAG TGCTTTGCAG TGAGTATCGT CAGAGTCCAG GAGCTGCTGT CCTGATCGCC	960
	GGCTGTGGCT TGAAGTTGAA GGTAGATTTG CGCCCGATCG GTTTGTATGC AGGTGTTGTG	1020
15	GTGGCCCCCG GCCTTGGGCG GCTCCCTGAT GTTGTGCGCT TCGCCGGCCG GCTTACCGAG	1080
	AAGAATTGGG GCCTTGGGCG TGAGCGGGCG GAGCAGCTCC GCCTCGCTGT TAGTGATTTG	1140
	CTCCGCAAGC TCACGAATGT AGCTCAGATG TGTGTGGATG TTGTTTCCCG TGTTTATGGG	1200
20	GTTCCTCCCTG GACTCGTTCA TAACCTGATT GGCATGCTAC AGGCTGTTGC TGATGGCAAG	1260
	GCACATTTCA CTGAGTCAGT AAAACCAGTG CTCGA	1295
25	<u>SEQ ID NO. 2:</u>	
	Arg Pro Val Pro Val Ala Ala Val Leu Pro Pro Cys Pro Glu Leu Glu	
	1 5 10 15	
30	Gln Gly Leu Leu Tyr Leu Pro Gln Glu Leu Thr Thr Cys Asp Ser Val	
	20 25 30	
	Val Thr Phe Glu Leu Thr Asp Ile Val His Cys Arg Met Ala Ala Pro	
	35 40 45	
35	Ser Gln Arg Lys Ala Val Leu Ser Thr Leu Val Gly Arg Tyr Gly Gly	
	50 55 60	
	Arg Thr Lys Leu Tyr Asn Ala Ser His Ser Asp Val Arg Asp Ser Leu	
40	65 70 75 80	
	Ala Arg Phe Ile Pro Ala Ile Gly Pro Val Gln Val Thr Thr Cys Glu	
	85 90 95	
45	Leu Tyr Glu Leu Val Glu Ala Met Val Glu Lys Gly Gln Asp Gly Ser	
	100 105 110	
	Ala Val Leu Glu Leu Asp Leu Cys Asn Arg Asp Val Ser Arg Ile Thr	
	115 120 125	
50	Phe Phe Gln Lys Asp Cys Asn Lys Phe Thr Thr Gly Glu Thr Ile Ala	
	130 135 140	
	His Gly Lys Val Gly Gln Gly Ile Ser Ala Trp Ser Lys Thr Phe Cys	
55	145 150 155 160	

	Leu Phe Gly Pro Trp Phe Arg Ala Ile Gly Ser Ala Ile Leu Ala
	165 170 175
5	Leu Leu Pro Gln Gly Val Phe Tyr Gly Asp Ala Phe Asp Asp Thr Val
	180 185 190
	Phe Ser Ala Ala Val Ala Ala Ala Lys Ala Ser Met Val Phe Glu Asn
	195 200 205
10	Asp Phe Ser Glu Phe Asp Ser Thr Gln Asn Asn Phe Ser Leu Gly Leu
	210 215 220
	Glu Cys Ala Ile Met Glu Glu Cys Gly Met Pro Gln Trp Leu Ile Arg
	225 230 235 240
15	Leu Tyr His Leu Ile Arg Ser Ala Trp Ile Leu Gln Ala Pro Lys Glu
	245 250 255
	Ser Leu Arg Gly Phe Trp Lys Lys His Ser Gly Glu Pro Gly Thr Leu
20	260 265 270
	Leu Trp Asn Thr Val Trp Asn Met Ala Val Ile Thr His Cys Tyr Asp
	275 280 285
25	Phe Arg Asp Phe Gln Val Ala Ala Phe Lys Gly Asp Asp Ser Ile Val
	290 295 300
	Leu Cys Ser Glu Tyr Arg Gln Ser Pro Gly Ala Ala Val Leu Ile Ala
	305 310 315 320
30	Gly Cys Gly Leu Lys Leu Lys Val Asp Phe Arg Pro Ile Gly Leu Tyr
	325 330 335
	Ala Gly Val Val Val Ala Pro Gly Leu Gly Ala Leu Pro Asp Val Val
35	340 345 350
	Arg Phe Ala Gly Arg Leu Thr Glu Lys Asn Trp Gly Pro Gly Pro Glu
	355 360 365
40	Arg Ala Glu Gln Leu Arg Leu Ala Val Ser Asp Phe Leu Arg Lys Leu
	370 375 380
	Thr Asn Val Ala Gln Met Cys Val Asp Val Val Ser Arg Val Tyr Gly
	385 390 395 400
45	Val Ser Pro Gly Leu Val His Asn Leu Ile Gly Met Leu Gln Ala Val
	405 410 415
	Ala Asp Gly Lys Ala His Phe Thr Glu Ser Val Lys Pro Val Leu
50	420 425 430

SEQ ID NO. 3:

55	Asp Leu Ser Leu Leu Gln Leu Phe Tyr His Pro Ala Pro Ser Ser Asn
	1 5 10 15

	Arg	Ala	Phe	Ser	Thr	Cys	Pro	Arg	Ser	Ser	Pro	Pro	Val	Ile	Val	Ser
				20					25					30		
5		His	Leu	Asn		Gln	Thr	Leu	Cys	Thr	Ala	Ala	Trp	Pro	Pro	Arg
			35					40					45			
	Ala	Ser	Ala	Arg	Pro	Cys	Cys	Pro	His	Ser	Trp	Ala	Ala	Thr	Ala	Val
			50				55						60			
10	Ala	Gln	Ser	Ser	Thr	Met	Leu	Pro	Thr	Leu	Met	Phe	Ala	Thr	Leu	Ser
		65				70					75					80
	Pro	Val	Leu	Ser	Arg	Pro	Leu	Ala	Pro	Tyr	Arg	Leu	Gln	Leu	Val	Asn
					85					90					95	
15		Cys	Thr	Ser		Trp	Arg	Pro	Trp	Ser	Arg	Arg	Ala	Arg	Met	Ala
					100					105					110	
	Pro	Ser	Leu	Ser	Leu	Ile	Phe	Ala	Thr	Val	Thr	Cys	Pro	Gly	Ser	Pro
20			115					120						125		
	Ser	Ser	Arg	Lys	Ile	Val	Thr	Ser	Ser	Pro	Gln	Val	Arg	Pro	Leu	Pro
			130				135						140			
25	Met	Val	Lys	Trp	Ala	Arg	Ala	Ser	Arg	Pro	Gly	Ala	Arg	Pro	Ser	Ala
		145				150					155					160
	Pro	Ser	Leu	Ala	Leu	Gly	Ser	Ala	Leu	Leu	Arg	Arg	Leu	Phe	Trp	Pro
				165						170					175	
30		Cys	Ser	Leu	Arg	Val	Cys	Phe	Thr	Val	Met	Pro	Leu	Met	Thr	Pro
				180						185					190	
	Ser	Arg	Arg	Leu	Trp	Pro	Gln	Gln	Arg	His	Pro	Trp	Cys	Leu	Arg	Met
35			195					200						205		
	Thr	Phe	Leu	Ser	Leu	Thr	Pro	Pro	Arg	Ile	Thr	Phe	Leu	Trp	Val	
		210						215						220		
40	Ser	Val	Leu	Leu	Trp	Arg	Ser	Val	Gly	Cys	Arg	Ser	Gly	Ser	Ser	Ala
		225				230					235					240
	Cys	Ile	Thr	Leu		Gly	Leu	Arg	Gly	Ser	Cys	Arg	Pro	Arg	Arg	Ser
				245						250					255	
45		Leu	Cys	Glu	Gly	Phe	Gly	Arg	Asn	Thr	Pro	Val	Ser	Pro	Ala	Leu
				260					265						270	
	Tyr	Gly	Ile	Leu	Ser	Gly	Ile	Trp	Pro	Leu	Leu	Pro	Thr	Val	Met	Thr
50			275					280						285		
	Ser	Ala	Ile	Phe	Arg	Trp	Leu	Pro	Leu	Lys	Val	Met	Ile	Arg		Cys
		290					295					300				
55	Phe	Ala	Val	Ser	Ile	Val	Arg	Val	Gln	Glu	Leu	Ser		Ser	Pro	
		305				310					315				320	

Ala Val Ala . Ser . Arg . Ile Ser Ala Arg Ser Val Cys Met  
 125 330 335

5 Gln Val Leu Trp Trp Pro Pro Ala Leu Ala Arg Ser Leu Met Leu Cys  
 340 345 350

Ala Ser Pro Ala Gly Leu Pro Arg Arg Ile Gly Ala Leu Ala Leu Ser  
 355 360 365

10 Gly Arg Ser Ser Ser Ala Ser Leu Leu Val Ile Ser Ser Ala Ser Ser  
 370 375 380

Arg Met . Leu Arg Cys Val Trp Met Leu Phe Pro Val Phe Met Gly  
 15 385 390 395 400

Phe Pro Leu Asp Ser Phe Ile Thr . Leu Ala Cys Tyr Arg Leu Leu  
 405 410 415

20 Leu Met Ala Arg His Ile Ser Leu Ser Gln . Asn Gln Cys Ser  
 420 425 430

SEQ ID NO. 4:

25 Thr Cys Pro Cys Cys Ser Cys Ser Thr Thr Leu Pro Arg Ala Arg Thr  
 1 5 10 15

Gly Pro Ser Leu Pro Ala Pro Gly Ala His His Leu . . Cys Arg  
 20 25 30

30 Asn Ile . Ile Asn Arg His Cys Ala Leu Pro His Gly Arg Pro Glu  
 35 40 45

Pro Ala Gln Gly Arg Ala Val His Thr Arg Gly Pro Leu Arg Arg Ser  
 35 50 55 60

His Lys Ala Leu Gln Cys Phe Pro Leu . Cys Ser Arg Leu Ser Arg  
 65 70 75 80

40 Pro Phe Tyr Pro Gly His Trp Pro Arg Thr Gly Tyr Asn Leu . Ile  
 85 90 95

Val Arg Ala Ser Gly Gly His Gly Arg Glu Gly Pro Gly Trp Leu Arg  
 100 105 110

45 Arg Pro . Ala . Ser Leu Gln Pro . Arg Val Gln Asp His Leu  
 115 120 125

Leu Pro Glu Arg Leu . Gln Val His His Arg . Asp His Cys Pro  
 50 130 135 140

Trp . Ser Gly Pro Gly His Leu Gly Leu Glu Gln Asp Leu Leu Arg  
 145 150 155 160

55 Pro Leu Trp Pro Leu Val Pro Arg Tyr . Glu Gly Tyr Ser Gly Pro  
 165 170 175

	Ala Pro Ser Gly Cys Val Leu Arg . Cys Leu . . His Arg Leu	180	185	190
5	Leu Gly Gly Cys Gly Arg Ser Lys Gly Ile His Gly Val . Glu .	195	200	205
	Leu Phe . Val . Leu His Pro Glu . Leu Phe Ser Gly Ser Arg	210	215	220
10	Val Cys Tyr Tyr Gly Gly Val Trp Asp Ala Ala Val Ala His Pro Pro	225	230	235
	Val Ser Pro Tyr Lys Val Cys Val Asp Leu Ala Gly Pro Glu Gly Val	245	250	255
	Ser Ala Arg Val Leu Glu Glu Thr Leu Arg . Ala Arg His Ser Ser	260	265	270
20	Met Glu Tyr Cys Leu Glu Tyr Gly Arg Tyr Tyr Pro Leu Leu . Leu	275	280	285
	Pro Arg Phe Ser Gly Gly Cys Leu . Arg . . Phe Asp Ser Ala	290	295	300
25	Leu Gln . Val Ser Ser Glu Ser Arg Ser Cys Cys Pro Asp Arg Arg	305	310	315
	Leu Trp Leu Glu Val Glu Gly Arg Phe Pro Pro Asp Arg Phe Val Cys	325	330	335
	Arg Cys Cys Gly Gly Pro Arg Pro Trp Arg Ala Pro . Cys Cys Ala	340	345	350
35	Leu Arg Arg Pro Ala Tyr Arg Glu Glu Leu Gly Pro Trp Pro . Ala	355	360	365
	Gly Gly Ala Ala Pro Pro Arg Cys . . Phe Pro Pro Gln Ala His	370	375	380
40	Glu Cys Ser Ser Asp Val Cys Gly Cys Cys Phe Pro Cys Leu Trp Gly	385	390	395
	Phe Pro Trp Thr Arg Ser . Pro Asp Trp His Ala Thr Gly Cys Cys	405	410	415
45	. Trp Gln Gly Thr Phe His . Val Ser Lys Thr Ser Ala Arg	420	425	430

50           The complementary strand, referred to here as the "reverse sequence," is set forth below in the same manner as the forward sequence set forth above. Several open reading frames, shorter than the long open reading frame found in the forward sequence, can

be seen in this reverse sequence. Because of the relative brevity of the open reading frames in the reverse direction, they are probably not expressed.

The following gene sequence has SEQ ID NO.5.

5 Reverse Sequence

SEQ ID NO. 5:

	TCGAGCACTG GTTTTACTGA CTCATGAAA TGTGCTTGG CATCAGCAAC AGCCTGTAGC	60
10	ATGCCAATCA GGTATGAA CAGTCAAGG GAAACCCCAT AAACACGGGA AACCAATCC	120
	ACACACATCT GAGCTACATT CGTGAGCTT CGGAGGAAAT CACTAACAGC GAGGCGGAGC	180
	TGCTCGGCC GCTCAGGGCC AGGCCCCAA TTCTTCTCG TAAGCCGGCC GGCGAAGCGC	240
15	ACAACATCAG GGAGCGCGCC AAGGCGGGG GCCACCACAA CACCTGCATA CAAACCGATC	300
	GGGCGGAAAT CTACCTTCAA CTTCAGGCA CAGCCGGCGA TCAGGACAGC AGCTCCTGGA	360
20	CTCTGACGAT ACTCACTGCA AAGCACTATC GAATCATCAC CTTTAAAGGC AGCCACCTGA	420
	AAATCGCGGA AGTCATAACA GTGGTAATA ACGGCCATAT TCCAGACAGT ATTCCATAGA	480
	AGAGTGCCGG GCTCAGCGGA GTGTTCTTC CAAACCCCTC GCAGAGACTC CTTCGGGGCC	540
25	TGCAAGATCC ACGCAGACCT TATAAGGTGA TACAGGCGGA TGAGCCACTG CGGCATCCCA	600
	CACTCCTCCA TAATAGCACA CTCTAGACCC AGAGAAAAGT TATTCTGGGT GGAGTCAAAC	660
30	TCAGAAAAGT CATTCTCAAA CACCATGGAT GCTTTTGCTG CGGCCACAGC CGCCGAGAAG	720
	ACGGTGTCAT CAAAGGCATC ACCGTAAAC ACACCCTGAG GGAGCAGGGC CAGAATAGCC	780
	TTCTCAATAG CGCGGAACCA AGGCCCCAAG AGGGCGCAGA AGGTCTTGCT CCAGGCCGAG	840
35	ATGCCCTGGC CCACTTTACC ATGGGCAATG GTCTCACCTG TGGTGAACCT GTTACAATCT	900
	TTCTGGAAGA AGGTGATCCT GGACACGTCA CGGTTGCAAA GATCAAGCTC AAGGACGGCG	960
40	GAGCCATCCT GGCCCTTCTC GACCATGGCC TCCACTAGCT CGTACAATTC ACAAGTTGTA	1020
	ACCTGTACGG GGCCAATGGC CGGGATAAAA CGGGCGAGAG AGTCGCGAAC ATCAGAGTGG	1080
	GAAGCATTGT AGAGCTTTGT CGGACCGCGG TAGCGGCCCA CGAGTGTGGA CAGCACGGCC	1140
45	TTGCGCTGGC TCGGGGCGGC CATGCGGCAG TGCACAATGT CTGTTAATTC AAATGTTACG	1200
	ACACTATCAC AGGTGGTGGC CTCTGGGGC AGGTAGAGAA GGCCCTGTTT GAGCTCGGGG	1260
50	CAGGGTGGTA GAACAGCTGC AACAGGGACA GGTCT	1295

Identity of this sequence with sequences in etiologic agents has been confirmed by locating a

corresponding sequence in a viral strain isolated in  
Burma. The Burmese isolate contains the following  
sequence of nucleotides (one strand and open reading  
frames shown). The following gene sequence has SEQ ID  
5 NO.6; the protein sequence corresponding to ORF1 has  
SEQ ID NO.7; ORF2 has SEQ ID NO.8; and ORF3 has SEQ ID  
NO.9.

SEQUENCE OF HEV (BURMA STRAIN)  
-ORF1-->

```

10      M E A H Q F I K A P G
      AGGCAGACCATATGTGTCGATGCCATGGAGGCCCATCAGTTTATTAAGGCTCCTGGC

      I T T A I E Q A A L A A A N S A L A N A
15  ATCACTACTGCTATTGAGCAGGCTGCTCTAGCAGCGGCCAACTCTGCCCTGGCGAATGCT 120

      V V V R P F L S H Q Q I E I L I N L M Q
      GTGGTAGTTAGGCCTTTCTCTCTCACCAGCAGATTGAGATCCTCATTAACTAATGCAA

20      P R Q L V F R P E V F W N H P I Q R V I
      CCTCGCCAGCTTGTCTTCGCGCCGAGGTTTCTGGAATCATCCATCCAGCGTGCATC 240

      H N E L E L Y C R A R S G R C L E I G A
      CATAACGAGCTGGAGCTTTACTGCCGCGCCGCTCCGGCCGCTGTCTTGAAATTGGCGCC

25      H P R S I N D N P N V V H R C F L R P V
      CATCCCCGCTCAATAAATGATAATCCTAATGTGGTCCACCGCTGCTTCTCCGCCCTGTT 360

      G R D V Q R W Y T A P T R G P A A N C R
30  GGGCGTGATGTTGAGCGCTGGTATACTGCTCCCACTCGCGGGCCGGCTGCTAATTGCCGG

      R S A L R G L P A A D R T Y C L D G F S
      CGTTCGCGCTGCGCGGGCTTCCCGCTGCTGACCGCACTTACTGCCTCGACGGGTTTTCT 480

35  G C N F P A E T G I A L Y S L H D M S P
      GGCTGTAACCTTCCCGCCGAGACTGGCATCGCCCTCTACTCCCTTCATGATATGTCACCA

      S D V A E A M F R H G M T R L Y A A L H
40  TCTGATGTCGCCGAGGCCATGTTCCGCCATGGTATGACGCGGCTCTATGCCGCCCTCCAT 600

      L P P E V L L P P G T Y R T A S Y L L I
      CTCCGCTGAGGTCTGCTGCCCGCTGGCAGATATCGCACCAGCATCGTATTTGCTAATT

      H D G R R V V V T Y E G D T S A G Y N H
45  CATGACGGTAGGCGCGTTGTGGTGACGTATGAGGGTGATACTAGTGCTGGTTACAACCAC 720

      D V S N L R S W I R T T K V T G D H P L
      GATGTCTCCAACCTTGGGCTCCTGGATTAGAACCACCAAGGTTACCGGAGACCATCCCCCTC

50  V I E R V R A I G C H F V L L L T A A P
      GTTATCGAGCGGGTTAGGGCCATTGGCTGCCACTTTGTCTCTTGCTCAGGCGAGCCCCG 840

      E P S P M P Y V P P P R S T E V Y V R S
55  GAGGCATCACCTATGCCTTATGTTCTTACCCCGGCTCTACCGAGGTCTATGTCCGATCG

```

1 F G P G B T P S L F P T S C S T K S T  
 ATCTTCGGGCGGGTGGGACCGCTTCTTATTCGCAAGCTCATGCTCCACTAAGTCGACC 960  
 5 F H A V P A H I W D R L M L F G A T L D  
 TTCCATGCTGTCCCTGGGCAATATTTGGGACGCTATTATGCTGTTCGGGCGCACCTTGGAT  
 D Q A F C C S R L M T Y L R G I S Y K V  
 GACCAAGCCTTTTGTGCTCCGCTTAAATGACCTACCTTCGGGCGATTAGCTACAAGGTC 1080  
 10 T V S T L V A N E G W N A S E D A L T A  
 ACTGTTGGTACCCCTGTGGCTAATGAAGGCTGGAATGCTCTGAGGACGCGCTCACAGCT  
 V I T A A Y L T I C H Q R L R T Q A I  
 GTTATCACTGCCGCTACCTTACCATTTGGCACCGAGGGTATCTCCGACCCAGGGCTATA 1200  
 15 S K G M R R L E R E H A Q K F I T R L Y  
 TCCAAGGGGATGCGTCTGTCTGGAACGGGACATGCCGAGAAGTTTATAACACGCGCTCTAC  
 S W L F E K S G R D Y I P G R Q L E F Y  
 AGCTGGCTCTTCGAGAAGTCCGGCGTGAATACATCCCTGGCGCTCAGTTGGAGTTCTAC 1320  
 20 A Q C R R W L S A G F H L D P R V L V F  
 GCCCAGTGCCAGGGCTGGGCTCTCCGCGGCTTTCATCTTGATCCACGGGTGTTGGTTTTT  
 D E S A P C H C R T A I R K A L S K F C  
 GACGAGTCGGCCCCCTGCCATTGTAGGACCGCGATCCGTAAGGCGCTCTCAAAGTTTTGC 1440  
 25 C F M K W L G Q E C T C F L Q P A E G A  
 TGCTTCATGAAGTGGCTTGGTCAGGAGTGCACTGCTTCTCAGCCTGCAGAAGGCGCC  
 30 V G D Q G H D N E A Y E G S D V D P A E  
 GTCGGCGACCGGGTCATGATAATGAAGCCTATGAGGGGTCCGATGTTGACCCTGTGTAG 1560  
 35 S A I S D I S G S Y V V P G T A L Q P L  
 TCCGCCATTAGTGACATATCTGGGTCTATGTCGTCCCTGGCACTGCCCTCCAACCGCTC  
 Y Q A L D L P A E I V A R A G R L T A T  
 TACCAAGGCCCTCGATCTCCCCGCTGAGATTGTGGCTCGCGCGGGCGGCTGACCGCCACA 1680  
 40 V K V S Q V D G R I D C E T L L G N K T  
 GTAAAGGTCTCCCAAGTCTGATGGGCGGATCGATTGCGAGACCCTTCTTGGTAACAAAACC  
 F R T S F V D G A V L E T N G P E R H N  
 TTTGCGACGTCGTTCTGTTGACGGGGCGGTCTTAGAGACCAATGGCCAGAGCGCCACAAT 1800  
 45 L S F D A S Q S T M A A G P F S L T Y A  
 CTCTCCTTCGATGCCAGTCAGAGCACTATGGCCGCTGGCCCTTTCAGTCTCACCTATGCC  
 A S A A G L E V R Y V A A G L D H R A V  
 GCCTCTGCAGCTGGGCTGGAGGTGCGCTATGTTGCTGCGGGGCTTGACCATCGGGCGGTT 1920  
 50 F A P G V S P R S A P G E V T A F C S A  
 TTTGCCCGCGGTGTTTACCCCGGTGAGCCCGGGGAGGTTACCGCCTTCTGCTCTGCC  
 L Y R F N R E A Q R H S L I G N L W F H  
 CTATACAGGTTTAAACGCTGAGGCGCCAGCGCCATTGCGTGATCGGTAACCTATGGTTCCAT 2040  
 55 P E G L I G L F A P F S P G H V W E S A  
 CCTGAGGGACTCATTTGGGCTCTTCGCGCGGCTTTTCGCGCGGGCATGTTTGGGAGTCGGCT



V C E L I R G A Y P M I Q T T S R V L R  
 GTATGCGAGGTCAATCGGTGGTGGATAACCGCATGATCCAGACCACTAGCCGGGTTCTCCGT 3360  
 5 S L F W G E P A Y G Q K L V F T Q A A K  
 TCGTTGTTCTGGGGTGAAGCGTGGCGTGGGGGAGAACTAGTGTTCACCCAGGCGGGCAAG  
 P A N P G S V T V H E A Q G A T Y T E T  
 CCGGCCAAACCGCGGTCAATGACGGTGGACGAGGGCGGCGGTACCTACACGGAGACC 3480  
 10 T I I A T A D A R G L I Q S S R A H A I  
 ACTATTATTGCCACAGCAGATGCGCGGGGGCTTATTGAGTGGTCTCGGGCTCATGCCATT  
 V A L T R H T E K D V I I D A P G L L R  
 GTTGCTCTGACGGCGCCACATGAGAAGTGGGTCAATGATGACGACCGAGGCTGCTTCGC 3600  
 15 E V G I S D A I V N N F F L A G G E I G  
 GAGGTGGGCATCTCCGATGGAAATCGTTAATAAGTTTTCTCGGTGGTGGCGAAATTGGT  
 H Q R P S Y I P R G N P D A N V D T L A  
 20 CACCAGCGCCCATCAGTTATCCCGTGGCAACCGTGAAGCAATGTTGACACCGCTGGCT 3720  
 A F P P S C Q I S A F H Q L A E E L G H  
 GCCTTCCCGCGGTCTTGCCAGATTAGTGGCTTCCATCAGTTGGCTGAGGAGCTTGGCCAC  
 25 R P V P V A A V L P P C P E L E Q G L L  
 AGACCTGTCCCTGTTGCAGCTGTTCTACCAACCGTGGCGCGAGCTCGAACAGGGCCTTCTC 3840  
 Y L P Q E L T T C D S V V T F E L T D I  
 30 TACCTGCCCCAGGAGCTCACCACCTGTGATAGTGTGGTAACATTTGAATTAACAGACATT  
 V H C R M A A P S Q R K A V L S T L V G  
 GTGCACTGCGCATGGCGCGCCCGAGCCAGCGCAAGGCGGTGCTGTCCACACTCGTGGGC 3960  
 35 R Y G G R T K L Y N A S H S D V R D S L  
 CGCTACGGCGGTGCGACAAAGCTCTACAATGCTTCCACTCTGATGTTCCGCACTCTCTC  
 A R F I P A I G P V Q V T T C E L Y E L  
 40 GCCCGTTTTATCCCGGCCATTGGCCCCGTACAGGTACAACTTGTGAATTGTACGAGCTA 4080  
 V E A M V E K G Q D G S A V L E L D L C  
 GTGGAGGCCATGGTCGAGAAGGCCAGGATGGCTCCGCGCTCCTTGAGCTTGATCTTTGC  
 N R D V S R I T F F Q K D C N K F T T G  
 45 AACCGTGACGTGTCCAGGATCACCTTCTTCCAGAAAGATTGTAACAAGTTCACCCAGGT 4200  
 E T I A H G K V G Q G I S A W S K T F C  
 GAGACCATTGCCCATGGTAAAGTGGGCGAGGCGATCTCGGCTGGAGCAAGACCTTCTGC  
 A L F G P W F R A I E K A I L A L L P Q  
 50 GCCCTCTTTGGCCCTTGGTTCCGCGCTATTGAGAAGGCTATTCTGGCCCTGCTCCCTCAG 4320  
 G V F Y G D A F D D T V F S A A V A A A  
 GGTGTGTTTACGGTGTGCTTTGATGACACCGTCTTCTCGGCGGCTGTGGCGCAGCA  
 55 K A S M V F E N D F S E F D S T Q N N F  
 AAGGCATCCATGGTGTGTTGAGATGACTTTTCTGAGTTTGACTCCACCCAGAATAACTTT 4440  
 S L G L E C A I M E E C G M P Q W L I R  
 60 TCTCTGGGTCTAGAGTGTGGTATTATGGAGGAGTGTGGGATGGCGAGTGGCTCATCCGC

L Y H L I R S A W I L Q A P K E S L R G  
 CTGTATCACCTTATAAGGTCTGCGTGGATCTTGCAGGCCCGAAGGAGTCTCTGCGAGGG 4560  
 5 F W K K H S G E P G T L L W N T V W N M  
 TTTTGGAAGAAACACTCCGGTGAGCCCGCACTCTTCTATGGAATACTGTCTGGAATATG  
 A V I T H C Y D F R D F Q V A A F K G D  
 GCCGTATTACCCACTGTTATGACTTCCGCGATTTTCAGGTGGCTGCCTTTAAAGGTGAT 4680  
 10 D S I V L C S E Y R Q S P G A A V L I A  
 GATTCGATAGTGCTTTGCAGTGAGTATCGTCAGAGTCCAGGAGCTGCTGCTCGATCGCC  
 G C G L K L K V D F R P I G L Y A G V V  
 GGCTGTGGCTTGAAGTTGAAGGTAGATTTCCGCCCGATCGGTTTGTATGCAGGTGTTGTG 4800  
 15 V A P G L G A L P D V V R F A G R L T E  
 GTGGCCCCGGCCTTGGCGCGCTCCCTGATGTTGTGCGCTTCGCCGGCCGGCTTACCGAG  
 K N W G P G P E R A E Q L R L A V S D F  
 20 AAGAATTGGGGCCCTGGCCCTGAGCGGGCGGAGCAGCTCCGCCCTCGCTGTTAGTGATTTTC 4920  
 L R K L T N V A Q M C V D V V S R V Y G  
 CTCCGCAAGCTCACGAATGTAGCTCAGATGTGTGGATGTTGTTTCCCGTGTATATGGG  
 25 V S P G L V H N L I G M L Q A V A D G K  
 GTTTCCTCGGACTCGTTCATAACCTGATTGGCATGCTACAGGCTGTTGCTGATGGCAAG 5040  
 A H F T E S V K P V L D L T N S I L C R  
 30 GCACATTTCACTGAGTCAGTAAAACCACTGCTCGACTTGACAAATTCAATCTGTGTGCG  
 | -ORF3--->  
 M N N M S F A A P M G S R P C A L G  
 35 V E Z M R P R P  
 GTGGAATGAATAACATGTCTTTTGTGCGCCCATGGGTTGCGGACCATGCGCCCTCGGCC 5160  
 | -ORF2-->  
 L F C C C S S C F C L C C P R H R P V S  
 40 I L L L L L M F L P M L P A P P P G Q P  
 TATTTTGTGCTGCTCCTCATGTTTTTGCCTATGCTGCCCGCGCCACCGCCCGGTACGCC  
 R L A A V V G G A A A V P A V V S G V T  
 45 S G R R R G R R S G G S G G G F W G D R  
 GTCTGGCCCGCGTCTGTTGGCGGCGCAGCGGCGGTTCCGGCGGTGTTTCTGGGGTGACCG 5280  
 G L I L S P S Q S P I F I Q P T P S P P  
 50 V D S Q P F A I P Y I H P T N P F A P D  
 GGTGATTCTCAGCCCTTCGCAATCCCTATATTCATCCAACCAACCCCTTCGCCCCCGA  
 M S P L R P G L D L V F A N P P D H S A  
 55 V T A A A G A G P R V R Q P A R P L G S  
 TGTCACCGCTGCGGCCGGGGCTGGACCTCGTGTTCCGCAACCCGCCGACCACTCGGCTC 5400



R G I A L T L F N L A D T L L G G L P T  
 CCGCGGGATAGCCCTCAGCCCTGTTCAACCTTGGTGACACTCTGCTTGGCGGCCCTGCCGAC  
 5 E L I S S A G G Q L F Y S R P V V S A N  
 AGAATTGATTTGGTGGGCTGGTGGCGAGCTGTTCTACTCCCGTCCCGTTGTCTCAGCCAA 6360  
 10 G E P T V K L Y T S V E N A Q Q D K G I  
 TGGCGAGCCGACTGTTAAGTTGTATACATCTGTAGAGAATGCTCAGCAGGATAAGGGTAT  
 A I P H D I D L G E S R V V I Q D Y D N  
 15 TGCAATCCCGCATGACATTGACCTCGGAGAACTCTGTTGGTTATTGAGGATTATGATAA 6480  
 Q H E Q D R P T P S P A P S R P F S V L  
 CCAACATGAACAAGATCGGCGGACGCTTCTCCAGCCCATCGCGCCCTTTCTCTGTCCT  
 20 R A N D V L W L S L T A A E Y D Q S T Y  
 TCGAGCTAATGATGTGCTTTGGCTCTCTCTCACCCTGCGGAGTATGACCACTCCACTTA 6600  
 25 G S S T G P V Y V S D S V T L V N V A T  
 TGGCTCTTCGACTGGCCAGTTTATGTTTCTGACTCTGTGACCTTGGTTAATGTTGCGAC  
 G A Q A V A R S L D W T K V T L D G R P  
 30 CGGCGCGCAGGCCGTTGCCCGGTGCTCGATTGGACCAAGGTCACACTTGACGGTCGCCC 6720  
 L S T I Q Q Y S K T F F V L P L R G K L  
 35 CCTCTCCACCATCCAGCAGTACTCGAAGACCTTCTTTGCTCCTGCCGCTCCGCGGTAAGCT  
 S F W E A G T T K A G Y P Y N Y N T T A  
 CTCTTTCTGGGAGGCAGGCACAATAAGCCGGGTACCTTATAATTATAACACCACTGC 6840  
 40 S D Q L L V E N A A G H R V A I S T Y T  
 TAGCGACCAACTGCTTGTGAGAATGCCGCCGGGCACCGGGTCGCTATTTCCACTTACAC  
 45 T S L G A G P V S I S A V A V L A P H S  
 CACTAGCCTGGGTGCTGGTCCCGTCTCCATTTCTGCGGTTGCCGTTTTAGCCCCCACTC 6960  
 A L A L L E D T L D Y P A R A H T F D D  
 50 TGGCTAGCATTGCTTGAGGATACCTTGGACTACCCTGCCCGCGCCCATACTTTTGATGA  
 F C P E C R P L G L Q G C A F Q S T V A  
 55 TTTCTGCCCAGAGTGCCGCCCCCTTGGCTTCAGGGCTGCGCTTTCCAGTCTACTGTGCG 7080  
 E L Q R L K M K V G K T R E L Z  
 60 TGAGCTTCAGCGCCTTAAGATGAAGGTGGGTAAAACTCGGGAGTTGTAGTTTATTGCTT

5

10

15

35

additional transcripts of 3.7 and 2.0 Kb were  
 identified using either of these epitopes as  
 hybridization probes. These polyadenylated  
 transcripts were identified using the extreme 3' end  
 5 epitope clone (406.3-2) as probe and therefore  
 established these transcripts as co-terminal with the  
 3' end of the genome (see below). One of the epitope  
 clones (406.4-2) was subsequently shown to react in a  
 specific fashion with antisera collected from 5  
 10 different geographic epidemics (Somalia, Burma,  
 Mexico, Tashkent and Pakistan). The 406.3-2 clone  
 reacted with sera from 4 out of these same 5 epidemics  
 (Yarborough et al., 1990). Both clones reacted with  
 only post inoculation antisera from infected cynos.  
 15 The latter experiment confirmed that seroconversion in  
 experimentally infected cynos was related to the  
 isolated exogenous cloned sequence.

A composite cDNA sequence (obtained from several  
 clones of the Mexican strain) is set forth below.

20 Composite Mexico strain sequence (SEQ ID NO.10):

SEQ ID NO. 10:

	GCCATGGAGG CCCACCAGTT CATTAAAGGCT CCTGGCATCA CTACTGCTAT TGAGCAAGCA	60
25	GCTCTAGCAG CGGCCAACTC CGCCCTTGCG AATGCTGTGG TGGTCCGGCC TTTCTTTTCC	120
	CATCAGCAGG TTGAGATCCT TATAAATCTC ATGCAACCTC GGCAGCTGGT GTTTCGTCCT	180
	GAGGTTTTTT GGAATCACCC GATTCAACGT GTTATACATA ATGAGCTTGA GCAGTATTGC	240
30	CGTGCTCGCT CGGGTCGCTG CCTTGAGATT GGAGCCCACC CACGCTCCAT TAATGATAAT	300
	CCTAATGTCC TCCATCGCTG CTTTCTCCAC CCGTCCGGCC GGGATGTTCA GCGCTGGTAC	360
35	ACAGCCCCGA CTAGGGGACC TGGCGGSAAC TGTCGCGGCT CGGCACTTCG TGGTCTGCCA	420
	CCAGCCGACC GCACTTACTG TTTTGATGGC TTTGCCGGCT GCCGTTTTGC CGCCGAGACT	480
	GGTGTGGCTC TCTATTCTCT CCATGACTTG CAGCCGGCTG ATGTTGCCGA GGCGATGGCT	540
40	CGCCACGGCA TGACCCGGCT TTATGCAGCT TTCCACTTGC CTCCAGAGGT GCTCCTGCCT	600
	CCTGGCACCT ACCGGAGATC ATCCTACTTG CTGATCCACG ATGGTAAGCG CGCGGTTGTC	660
45	ACTTATGAGG GTGACACTAG GCGCGGTTAC AATCATGATG TTGCCACCCT CCGCACATGG	720

	ATCAGGACAA CTAAGGTTGT GGGTGAACAC CTTTGGTGA TCGAGCGGGT GCGGGGTATT	780
	GGCTGTCACT TTGTGTTGTT GATGATGGB GCGCGTBAAC CTTCCCGGAT GCCCTACGTT	840
5	CCTTACCCGC GTTGGACGGA GGTGTATGTC GGTGTATCT TTGGGCGCGG CCGGTCCCCG	900
	TGGGTGTTCC CGACCGCTTG TGTTGTGAG TCACTTTTC AGCGCGTCCG CACGCACATC	960
10	TGGGACCGTC TCATGCTCTT TGGGCGGAC GTGACGACG AGGCGTTTTC CTGCTCCAGG	1020
	CTTATGACGT ACCTTGTGTC GATTAGCTAT GAGGTAAGTC TGGGTGCCCT GGTGCTAAT	1080
	GAAGGCTGGA ATGCCACCGA GATGCGCTC ACTGAGTTA TTACGGCGGC TTACCTCACA	1140
15	ATATGTGATC AGCGTTATTT GCGGACCGAG GCGATTTCTA AGGGCATGCG CCGGCTTGAG	1200
	CTTGAACATG CTCAGAAATT TATTTCAGGC GTGTACAGCT GGTATTTGA GAAGTCAGGT	1260
20	CGTGATTACA TCCGAGGCGG CCAGGTGAG TTCTACGCTC AGTGCGCGCG CTGGTTATCT	1320
	GCCGGGTTCC ATCTCGACCC CCGCACCTTA GTTTTTGATG AGTCAGTGCC TTGTAGCTGC	1380
	CGAACCACCA TCCGGCGGAT CGGTGAAAA TTTTGCTGTT TTATGAAGTG GCTCGGTCAG	1440
25	GAGTGTTCCT GTTTCTCCA GCCCGCGGAG GGGCTGGCGG GCGACCAAGG TCATGACAAT	1500
	GAGGCCTATG AAGGCTCTGA TGTGATACT GGTGAGCCTG CCACCTAGA CATTACAGGC	1560
30	TCATACATCG TGGATGGTCG GTCTCTGCAA ACTGTCTATC AAGCTCTCGA CCTGCCAGCT	1620
	GACCTGGTAG CTCGCGCAGC CCGACTGTCT GCTACAGTTA CTGTTACTGA AACCTCTGGC	1680
	CGTCTGGATT GCCAAACAAT GATCGGCAAT AAGACTTTTC TCACTACCTT TGTTGATGGG	1740
35	GCACGCCTTG AGGTTAACGG GCCTGAGCAG CTTAACCTCT CTTTGTACAG CCAGCAGTGT	1800
	AGTATGGCAG CCGGCCCGTT TTGCTCACC TATGCTGCCG TAGATGGCGG GCTGGAAGTT	1860
40	CATTTTTCCA CCGCTGGCTT CGAGAGCGCT GTTGTTTTCC CCGCTGGTAA TGCCCCGACT	1920
	GCCCCGCCGA GTGAGGTGAC CGCTTCTGTC TCAGCTCTTT ATAGGCACAA CCGGCAGAGC	1980
	CAGCGCCAGT CGGTTATTGG TAGTTGTGG CTGCACCTG AAGGTTTGCT CGGCCTGTTT	2040
45	CCGCCCTTTT CACCCGGGCA TGAGTGGCGG TCTGCTAACC CATTTTGCGG CGAGAGCACG	2100
	CTCTACACCC GCACTTGGTC CACAATTACA GACACACCTT TAACTGTGCG GCTAATTTCC	2160
50	GGTCATTGCG ATGCTGCTCC CCACTCGGGG GGGCCACCTG CTACTGCCAC AGGCCCTGCT	2220
	GTAGGCTCGT CTGACTCTCC AGACCTGAC CCGCTACCTG ATGTTACAGA TGGCTCACGC	2280
	CCCTCTGGGG CCGTCCGGT TGGCGGACG CGGAATGGCG TTCCGCAGCG CCGCTTACTA	2340
55	CACACCTACC CTGACGGGCG TAAGATGAT GTGGGTCCA TTTTCGAGTC TGAGTGCACC	2400

	TGGCTTGTC	ACGATGTAA	CGCGGCGAC	CGCCCTGGT	GCGGGCTTTG	TCATGCTTTT	2460
	TTTCAGCGT	ACCGTGAAT	GTTCAGCGG	ACCAAGTTTG	TGATGCGTGA	TGGTCTTGCC	2520
5	GCGTATACCC	TTACACCGG	GCGATGATT	CATGCGGTGG	CCCCGGACTA	TCGATTGGAA	2580
	CATAACCCCA	AGAGGCTGA	GGTGGCTAC	CGCGAGACTT	GCSCCCGCG	AGGCACTGCT	2640
10	GCCTATCCAC	TCTTAGCGG	TGGCATTTAC	CAGGTGCCTG	TTAGTTTGAG	TTTTGATGCC	2700
	TGGGAGCGGA	ACCAACCGG	GTTCAGCGG	CTTTACCTAA	CAGAGCTGGC	GGCTCGGTGG	2760
	TTTGAATCCA	ACCGCGCGG	TCAGCGCGG	TTGAACATAA	CTGAGGATAC	CGCCCGTGCG	2820
15	GCCAACCTGG	CGCTGGAGT	TGACTCGGG	AGTGAAGTAG	GCGCGCATG	TGCCGGGTGT	2880
	AAAGTCGAGC	CTGGCGTTG	CGGTATCAG	TTTACAGCG	GTGTCCCGG	CTCTGGCAAG	2940
20	TCAAAGTCCG	TGCAACAGG	GGATGTGGAT	GTTGTTGTTG	TGCCACTCG	CGAGCTTCGG	3000
	AACGCTTGGC	GGCGCGGGG	CTTGGCGCA	TTCACTCCGC	ACACTGCGGC	CCGTGTCACT	3060
	AGCGGCGTA	GGTTGTGAT	TGATGAGGCC	CCTTCGCTCC	CCCCACACTT	GCTGCTTTTA	3120
25	CATATGCAGC	GTGCTGCATC	TGTGCACCTC	CTTGGGGACC	CGAATCAGAT	CCCCGCCATA	3180
	GATTTTGAGC	ACACCGGTCT	GATTCCAGCA	ATACGGCCGG	AGTTGGTCCC	GACTTCATGG	3240
30	TGGCATGTCA	CCACCGTTG	CCCTGCAGAT	GTCTGTGAGT	TAGTCCGTGG	TGCTTACCCT	3300
	AAAATCCAGA	CTACAAGTAA	GGTGCTCCGT	TCCCTTTTCT	GGGGAGAGCC	AGCTGTGCGC	3360
	CAGAAGCTAG	TGTTACACA	GGCTGCTAAG	GCCGCGCACC	CCGGATCTAT	AACGGTCCAT	3420
35	GAGGCCCAGG	GTGCCACTTT	TACCACTACA	ACTATAATTG	CAACTGCAGA	TGCCCGTGGC	3480
	CTCATACAGT	CCTCCCGGGC	TCACGCTATA	GTTGCTCTCA	CTAGGCATAC	TGAAAAATGT	3540
40	GTTATACTTG	ACTCTCCCGG	CCTGTTGCGT	GAGGTGGGTA	TCTCAGATGC	CATTGTTAAT	3600
	AATTTCTTCC	TTTCGGGTGG	CGAGGTGGT	CACCAGAGAC	CATCGGTCAT	TCCGCGAGGC	3660
	AACCTGACC	GCAATGTTGA	CGTGCTTGGC	GCGTTTCCAC	CTTCATGCCA	AATAAGCGCC	3720
45	TTCCATCAGC	TTGCTGAGGA	GCTGGGCGAC	CGGCGGGCGC	CGGTGGCGGC	TGTGCTACCT	3780
	CCCTGCCCTG	AGCTTGAGCA	GCGCTTCTC	TATCTGCCAC	AGGAGCTAGC	CTCCTGTGAC	3840
50	AGTGTGTGA	CATTTGAGCT	AACTGACATT	GTGCACTGCC	GCATGGCGGC	CCCTAGCCAA	3900
	AGGAAAGCTG	TTTGTCCAC	GCTGGTAGGC	CGTATGGCA	GACGCACAAG	GCTTTATGAT	3960
	GCGGGTCACA	CCGATGTCCG	CGCTCCCTT	GCGCGCTTTA	TTCCCACTCT	CGGGCGGGTT	4020
55	ACTGCCACCA	CGTGTGAGT	CTTTGAGCTT	GTAGAGGCGA	TGGTGGAGAA	GGGCCAAGAC	4080

	GGTTCAGCCG TCCTGAGTT GGAATTTGTC AGCGAGATG TGTCCCGCAT AACCTTTTTC	4140
	CAGAAGGATT GTAACAGTT CAGACGCGC GAGACATTC CGCATGGCAA AGTCGGTCAG	4200
5	GGTATCTTCC GCTGGAGTAA GAGTTTTGT GCGCTGTTG GCGCCTGGTT CCGTGCGATT	4260
	GAGAAGGCTA TTCTATCGCT TTTACGACA GCTGTCTCT ACGGGGATGC TTATGACGAC	4320
10	TCAGTATTCT CTGCTGCGT GGTGGGCGC AGCATGCGA TGGTGTGTA AAATGATTTT	4380
	TCTGAGTTTG ACTCGACTCA GAATACTTT TCGTAGGTG TTGAGTGGC CATTATGGAA	4440
	GAGTGTGGTA TGCCCCAGT GCTTGTGAG TTGTACCAT GCGTCCGGTC GCGGTGGATC	4500
15	CTGCAGGCCC CAAAAGAGTC TTGAGAGGG TTCTGGAAG AGCATTCTGG TGAGCCGGGC	4560
	AGCTTGCTCT GGAATACGCT GTGGAAGAT GCAATCATTC CCGATTGCTA TGAGTCCGG	4620
20	GACCTCCAGG TTGCGGCTT CAAGGGCGAC GACTCGGTG TCCTCTGTAG TGAATACCGC	4680
	CAGAGCCCAG GCGCCGGTTC GCTTATAGCA GGCTGTGGTT TGAAGTTGAA GGCTGACTTC	4740
	CGGCCGATTG GGCTGTATGC CGGGGTTGTC GTCGCCCCGG GGCTCGGGGC CCTACCCGAT	4800
25	GTCGTTGAT TCGCCGACG GCTTTCGGAG AAGAACTGGG GGCTGATCC GGAGCGGGCA	4860
	GAGCAGCTCC GCCTCGCGT GCAGGATTTT CTCTAGGT TAACGAATGT GGCCAGATT	4920
30	TGTGTTGAGG TGGTGTCTAG AGTTTACGGG GTTCCCGG GTCTGGTTCA TAACCTGATA	4980
	GGCATGCTCC AGACTATTGG TGATGGTAAG GCGCATTTTA CAGAGTCTGT TAAGCCTATA	5040
	CTTGACCTTA CACACTCAAT TATGCACCG TCTGAATGAA TAACATGTGG TTTGCTGCGC	5100
35	CCATGGGTTT GCCACCATGC GCGTAGGCG TCTTTGCTG TTGTTCTCT TGTTCCTGCC	5160
	TATGTTGCC GCGCCACGA CCGGTACGC GTCTGGCGC CGTCGTGGGC GGCGCAGCGG	5220
40	CGGTACCGGC GGTGGTTTCT GGGGTGACG GGTGATTCT CAGCCCTTCG CAATCCCCTA	5280
	TATTCATCCA ACCAACCCT TTGCCCCAGA CGTTGCCGT GCGTCCGGT CTGGACCTCG	5340
	CCTTCGCCAA CCAGCCCGC CACTTGGTC CACTTGGCA GATCAGGCC AGCGCCCTC	5400
45	CGCTGCCTCC CGTCGCGAC CTGCCACAG CGGGGTGCG GCGCTGACG CTGTGGCGCC	5460
	TGCCCATGAC ACCTACCGG TCCGGACGT TGATTCTGC GGTGCAATTC TACGCCCA	5520
50	GTATAATTTG TCTACTTAC CCGTACATC CTCTGTGGC TGTGGCACTA ATTTAGTCT	5580
	GTATGCAGCC CCGCTTAAT CCGCTCTGC GCTGCAGGAC GGTACTAATA CTCACATTAT	5640
	GGCCACAGAG GCCTCCAAT ATGCACGTA CCGGTTGCC CGCGCTACTA TCCGTTACCG	5700
55	GCCCTAGTG CCAATGCAG TTGAGGCTA TGTATATCC ATTTCTTCT GGCTCAAAC	5760

	AACCACAACC CCTACATCTG TTGACATGAA TTCCATTACT TCCACTGATG TCAGGATTCT	5820
	TGTTCAACCT GGCATAGCAT CTGAATTGGT CATCCCAAGC GAGCGCCTTC ACTACCGCAA	5880
5	TCAAGGTTGG CGCTCGGTG AGACATCTGG TGTTCGTGAG GAGGAAGCCA CCTCCGGTCT	5940
	TGTCATGTTA TGCATACATG GCTCTCCAGT TAACTCCTAT ACCAATACCC CTTATACCGG	6000
10	TGCCCTTGGC TTACTGGAGT TTGCTTAGA GOTTGAGTTT CGCAATCTCA CCACCTGTAA	6060
	CACCAATACA GGTGTGTCCC GTTACTCCAG CACTGCTCGT CACTCCGCCC GAGGGGCCGA	6120
	CGGGACTGCG GAGCTGACCA CAACTGCAGC CACCAGGTTG ATGAAAGATC TCCACTTTAC	6180
15	CGGCCTTAAT GGGGTAGGTG AAGTGGGCGG CGGGATAGCT CTAACATTAC TTAACCTTGC	6240
	TGACACGGTC CTCGGCGGGC TCCGACAGA ATTAATTTGG TCGGCTGGCG GGCAACTGTT	6300
20	TTATTCCCGC CCGGTTGTCT CAGCCAATGG CGAGCCAACC GTGAAGCTCT ATACATCAGT	6360
	GGAGAATGCT CAGCAGGATA AGGGTGTTCG TATCCCCCAC GATATCGATC TTGGTGATTC	6420
	GCGTGTGGTC ATTCAAGGAT ATGACAACCA GCATGAGCAG GATCGGCCCA CCCCCTCGCC	6480
25	TGCGCCATCT CGGCCTTTTT CTGTTCTCGG AGCAAATGAT GTACTTTGGC TGTCCCTCAC	6540
	TGCAGCCGAG TATGACCAGT CCACTTACGG GTCGTCAACT GGCCCGGTTT ATATCTCGGA	6600
	CAGCGTGA CT TTGGTGAATG TTGCGACTGG CGCGCAGGCC GTAGCCCGAT CGCTTGACTG	6660
30	GTCCAAAGTC ACCCTCGACG GCGGGCCCTT CCCGACTGTT GAGCAATATT CCAAGACATT	6720
	CTTTGTGCTC CCCCTTCGTG GCAAGCTCTC CTTTGGGAG GCCGGCACAA CAAAAGCAGG	6780
35	TTATCCTTAT AATTATAATA CTA CTGCTAG TGACCAGATT CTGATTGAAA ATGCTGCCGG	6840
	CCATCGGGTC GCCATTTCAA CCTATACCAC CAGGCTTGGG GCCGGTCCGG TCGCCATTTT	6900
40	TGCGGCCGCG GTTTTGGCTC CACGCTCCGC CCTGGCTCTG CTGGAGGATA CTTTTGATTA	6960
	TCCGGGGCGG GCGCACACAT TTGATGACTT CTGCCCTGAA TGCCGCGCTT TAGGCCTCCA	7020
	GGGTTGTGCT TTCCAGTCAA CTGTGCTGA GCTCCAGCGC CTTAAAGTTA AGGTGGGTAA	7080
45	AACTCGGGAG TTGTAGTTTA TTTGGCTGTG CCCACCTACT TATATCTGCT GATTCCTTT	7140
	ATTTCTTTT TCTCGGTCCC GCGCTCCCTG A	7171

50           The above sequence was obtained from  
polyadenylated clones. For clarity the 3' polyA  
"tail" has been omitted.

The sequence above includes a partial cDNA sequence consisting of 1661 nucleotides that was identified in a previous application in this series. The previously identified partial sequence is set forth below, with certain corrections (SEQ ID NO.11). The corrections include deletion of the first 80 bases of the prior reported sequence, which are cloning artifacts; insertion of G after former position 174, of C after 270, and of GGCG after 279; change of C to T at former position 709, of GC to CG at 722-723, of CC to TT at 1238-39, and of C to G at 1606; deletion of T at former position 765; and deletion of the last 11 bases of the former sequence, which are part of a linker sequence and are not of viral origin.

15 Non-A Non-B T: Mexican Strain; SEQ ID NO.11  
 SEQ ID NO. 11:

	GTTGCGTGAG GTGGGATCT CAGATGCGAT TGTAAATAAT TTCTTCCTTT CGGGTGCGCA	60
20	GGTTGGTCAC CAGAGACCAT CGGTCATTCG GCGAGGCAAC CCTGACCGCA ATGTTGACGT	120
	GCTTGCGGCG TTTCACCTT CATGCCAAT AAGCGCTTC CATCAGCTTG CTGAGGAGCT	180
	GGGCCACCGG CCGGCGCGG TGGCGGCTGT GCTACCTCCC TGCCCTGAGC TTGAGCAGGG	240
25	CCTTCTCTAT CTGCCACAGG AGCTAGCCTC CTGTGACAGT GTTGTGACAT TTGAGCTAAC	300
	TGACATTGTG CACTGCCGCA TGGCGGCCCC TAGCCAAAGG AAAGCTGTTT TGTCCACGCT	360
30	GGTAGGCCGG TATGGCAGAC GCACAAGGCT TTATGATGCG GGTACACCG ATGTCCGCGC	420
	CTCCCTTGCG CGCTTTATTC CCACTCTCGG GCGGTTACT GCCACCACCT GTGAACTCTT	480
	TGAGCTTGTA GAGGCGATGG TGGAGAAAGG CCAAGACGGT TCAGCCGTCC TCGAGTTGGA	540
35	TTTGTGCAGC CGAGATGTCT CCGGCATAC CTTTTCCAG AAGGATTGTA ACAAGTTCAC	600
	GACCGGCGAG ACAATTGCGC ATGGCAAGT GGGTCAGGT ATCTTCGCT GGAGTAAGAC	660
40	CTTTGTGCG CTGTTTGCG CCTGGTTGCG TGGATTGAG AAGGCTATTC TATCCCTTTT	720
	ACCACAAGCT GTGTTTACG GGGATGTTA TGAGACTCA GTATTCTCTG CTGCCGTGGC	780
	TGGCGCCAGC CATGCCATGG TGTGAAAAA TGATTTTCT GAGTTTGACT CGACTCAGAA	840
45	TAACTTTTC CTAGGTCTT GGTGGCGAT TATGGAAGAG TGTGGTATGC CCCAGTGGCT	900
	TGTCAGGTTG TACCATGCGG TCGGTGCGG GTGGATCTG CAGGCCCCAA AAGAGTCTTT	960

	GAGAGGGTTC TGGAAAGAAGC ATTCTGGTGA GCGGGGACAG TTGCTCTGGA ATACGGTGTG	1020
	GAACATGGCA ATGATTGCCC ATTGCTATGA GTTCCGGGAC CTCGAGGTG CCGCCTTCAA	1080
5	GGGCGACGAC TCGGTGGTCC TGTGTAGTGA ATAGCGCCAG AGCCCAGGCG CCGGTTGGCT	1140
	TATAGCAGGC TGTGGTTTGA AGTTGAAGGC TGAATTCCGG CCGATTGGGC TGTATGCCGG	1200
10	GGTTGTGGTC GCGCGGGGGC TCGGGGCCCC ACCCGATGTC GTTCGATTTC CCGGACGGCT	1260
	TTCGGAGAAG AACTGGGGGC CTGATCGGGA GCGGGCAGAG CAGCTCCGCC TCGCCGTGCA	1320
	GGATTTCTTC CGTAGGTTAA GGAATGTGGC CCAGATTTGT GTTGAGGTGG TGTCTAGAGT	1380
15	TTACGGGGTT TCCCCGGGTC TGGTTGATAA CCGGATAGGC ATGCTCCAGA CTATTGGTGA	1440
	TGGTAAGGGC CATTTTACAG AGTGTGTTAA GCGTATACTT GACCTTACAC ACTCAATTAT	1500
	GCACCGGTCT GAATGAATAA CATGTGGTTT GCTGCGCCCA TGGGTTGCC ACCATGCGCC	1560
20	CTAGGCCTCT TTTCG	1575

25 When comparing the Burmese and Mexican strains, 75.7% identity is seen in a 7189 nucleotide overlap beginning at nucleotide 1 of the Mexican strain and nucleotide 25 of the Burmese strain.

30 In the same manner, a different strain of HEV was identified in an isolate obtained in Tashkent, U.S.S.R. The Tashkent sequence is given below (SEQ ID NO.12):

SEQ ID NO. 12:

35	CGGGCCCCGT ACAGGTCACA ACCTGTGAGT TGTACGAGCT AGTGGAGGCC ATGGTCGAGA	60
	AAGGCCAGGA TGGCTCCGCC GTCCTTGAGC TCGATCTCTG CAACCGTGAC GTGTCCAGGA	120
	TCACCTTTTT CCAGAAAGAT TGCAATAAGT TCACCACGGG AGAGACCATC GCCCATGGTA	180
40	AAGTGGGCCA GGGCATTTCG GCGTGGAGTA AGACCTTCTG TGCCCTTTTC GGCCCTGGT	240
	TCCGTGCTAT TGAGAAGGCT ATTCTGGCCC TGCTCCCTCA GGGTGTGTTT TATGGGGATG	300
45	CCTTTGATGA CACCGTCTTC TCGCGCGGTG TGGCCGCAGC AAAGGCGTCC ATGGTGTGTTG	360
	AGAATGACTT TTCTGAGTTT GACTCCACCC AGAATAATTT TTCCCTGGGC CTAGAGTGTG	420
	CTATTATGGA GAAGTGTGGG ATGCCGAAGT GGTCATCCG CTTGTACCAC CTTATAAGGT	480
50	CTGCGTGGAT CCGGAGGCC CCGAAGGAGT CCGTCCGAGG GTGTTGGAAG AAACACTCCG	540
	GTGAGCCCGG CACTCTTTCTA TGGATACTG TGTGGAACAT GGCCGTTATC ACCCATGTGTT	600

ACGATTTCCG CGATTTGCAG GTGGCTGCTT TTAAGGTGA TGATTGATA GTGCTTTGCA 660  
 GTGAGTACCG TCAGAGTCCA GGGGCTGCTG TCGTATTGC TGGCTGTGGC TTAAGCTGA 720  
 5 AGGTGGGTTT CCGTCCGATT GGTGTTATG CAGGTGTTGT GGTGACCCCC GGCCTTGGCG 780  
 CGCTTCCCGA CCGCTGGGCG TTGTCCGGCC GGCTTACTGA GAAGAATTGG GGCCTTGGCC 840  
 10 CTGAGCGGGC GGAGCAGCTC CGCCTTGCTG TGCG 874

As shown in the following comparison of sequences, the Tashkent (Tash.) sequence more closely resembles the Burma sequence than the Mexico sequence, as would be expected of two strains from more closely related geographical areas. The numbering system used in the comparison is based on the Burma sequence. As indicated previously, Burma has SEQ ID NO:6; Mexico, SEQ ID NO:10; and Tashkent, SEQ ID NO:12. The letters present in the lines between the sequences indicate conserved nucleotides.

		10v	20v	30v	40v	50v	60v
25	-BURMA	AGGCAGACCACATATGTGGT	CGATGCCATGGAGG	CCCCATCAGTTT	ATTAAGGCTCCTGGCA		
				GCCATGGAGG	CCCCA CAGTT ATTAAGGCTCCTGGCA		
	-MEXICO			GCCATGGAGG	CCCCACCAGTTCATTAAGGCTCCTGGCA		
		70v	80v	90v	100v	110v	120v
30	-BURMA	TCCTACTGCTATTGAGCAGGCTGCTCTAGCAGCGGCCAACTCTGCCCTGGCGAATGCTG					
		TCCTACTGCTATTGAGCA GC GCTCTAGCAGCGGCCAACTC GCCCT GCGAATGCTG					
	-MEXICO	TCCTACTGCTATTGAGCAAGCAGCTCTAGCAGCGGCCAACTCCGCCCTTGCGAATGCTG					
		130v	140v	150v	160v	170v	180v
35	-BURMA	TGGTAGTTAGGCTTTTCTCTCTCACCAGCAGATTGAGATCCTCATTAACCTAATGCAAC					
		TGGT GT GGCCTTT CT TC CA CAGCAG TTGAGATCCT AT AA CT ATGCAAC					
	-MEXICO	TGGTGGTCCGGCCTTTCTCTTCCCATCAGCAGGTTGAGATCCTTATAAATCTCATGCAAC					
		190v	200v	210v	220v	230v	240v
40	-BURMA	CTCGCCAGCTTGTGTTTCCGGCCCGAGGTTTTCTGGAATCATCCCATCCAGCGTGTCTATCC					
		CTCG CAGCT GT TT CG CC GAGGTTTT TGGAAATCA CC AT CA CGTGT AT C					
	-MEXICO	CTCGGCAGCTGGTGTGTTCTGCTGAGGTTTTTGGAAATCACCCGATTCAACGTGTTATAC					
		250v	260v	270v	280v	290v	300v
45	-BURMA	ATAACGAGCTGGAGCTTTACTGCCGCGCCCGCTCCGCGCGTGTCTTGAAATTGGCGCCC					
		ATAA GAGCT GAGC TA TGCCG GC CGCTC GG CGCTG CTTGA ATTGG GCCC					
	-MEXICO	ATAATGAGCTTGAGCAGTATTGCCGTGCTCGCTCGGGTCTGCTGCCTTGAGATTGGAGCCC					
		310v	320v	330v	340v	350v	360v
50	-BURMA	ATCCCGCTCAATAAATGATAATCCTAATGTGGTCCACCGCTGCTTCTCCGCCCTGTTG					
		A CC CGCTC AT AATGATAATCCTAATGT TCCA CGCTGCTT CTCC CCC GT G					
	-MEXICO	ATCCACGCTCCATTAATGATAATCCTAATGTCTCCATCGCTGCTTCTCCACCCCGCTG					

[illegible]

		1030v	1040v	1050v	1060v	1070v	1080v
	-BURMA	ACCAAGGCTTTTGTGCTTTTATGAGTACCTTGGCGGCATTAGCTACAAGGTCA					
		ACCA GCTTTTGTGCTTTTATGAGTACCTTGGCGGCATTAGCTA AAGGT A					
5	-MEXICO	ACCAAGGCTTTTGTGCTTTTATGAGTACCTTGGCGGCATTAGCTATAAGGTAA					
		1090v	1100v	1110v	1120v	1130v	1140v
	-BURMA	CTGTGCTTGTGCTTTTGTGCTTTTATGAGTACCTTGGCGGCATTAGCTACAAGGTCA					
		CTGT GCT GCTT GT GCTTTTGTGCTTTTATGAGTACCTTGGCGGCATTAGCTA AAGGT A					
10	-MEXICO	CTGTGCTTGTGCTTTTGTGCTTTTATGAGTACCTTGGCGGCATTAGCTACAAGGTCA					
		1150v	1160v	1170v	1180v	1190v	1200v
	-BURMA	TTATCACTGCTGCTTGTGCTTTTATGAGTACCTTGGCGGCATTAGCTACAAGGTCA					
		TTAT AC GC GC TACCT AC AT TG CA CAGCG TAT T CG ACCCAGGC AT T					
15	-MEXICO	TTATCACTGCTGCTTGTGCTTTTATGAGTACCTTGGCGGCATTAGCTACAAGGTCA					
		1210v	1220v	1230v	1240v	1250v	1260v
	-BURMA	CCAAGGGGATGCTGCTTGTGCTTTTATGAGTACCTTGGCGGCATTAGCTACAAGGTCA					
		C AAGGG ATGCG CG CT GA C BA CATGC CAGAA TTTAT CACGCTCTACA					
20	-MEXICO	CCAAGGGGATGCTGCTTGTGCTTTTATGAGTACCTTGGCGGCATTAGCTACAAGGTCA					
		1270v	1280v	1290v	1300v	1310v	1320v
	-BURMA	GCTGGCTTGTGCTTGTGCTTTTATGAGTACCTTGGCGGCATTAGCTACAAGGTCA					
		GCTGGCT TT GAGAAGTC GG CGTGATTACATCCC GGCCG CAG TG AGTTCTACG					
25	-MEXICO	GCTGGCTTGTGCTTGTGCTTTTATGAGTACCTTGGCGGCATTAGCTACAAGGTCA					
		1330v	1340v	1350v	1360v	1370v	1380v
	-BURMA	CCCAAGGGGATGCTGCTTGTGCTTTTATGAGTACCTTGGCGGCATTAGCTACAAGGTCA					
		C CAGTGC G CGTGG T TC GCCGG TT CATCT GA CC CG TT GTTTTG					
30	-MEXICO	CCCAAGGGGATGCTGCTTGTGCTTTTATGAGTACCTTGGCGGCATTAGCTACAAGGTCA					
		1390v	1400v	1410v	1420v	1430v	1440v
	-BURMA	ACGAGTGGGCTTGTGCTTGTGCTTTTATGAGTACCTTGGCGGCATTAGCTACAAGGTCA					
		A GAGTC G CC TG TG G ACC C ATCCG G AAA TTTTGCT					
35	-MEXICO	ACGAGTGGGCTTGTGCTTGTGCTTTTATGAGTACCTTGGCGGCATTAGCTACAAGGTCA					
		1450v	1460v	1470v	1480v	1490v	1500v
	-BURMA	GCTTCATGAAGTGGCTTGTGCTTGTGCTTTTATGAGTACCTTGGCGGCATTAGCTACAAGGTCA					
		G TT ATGAAGTGGCT GGTGAGGAGTG C TG TTCCT CAGCC GC GA GG G					
40	-MEXICO	GCTTCATGAAGTGGCTTGTGCTTGTGCTTTTATGAGTACCTTGGCGGCATTAGCTACAAGGTCA					
		1510v	1520v	1530v	1540v	1550v	1560v
	-BURMA	TCGGCGACCAAGGCTGATGATAATGAAGCTATGAGGGGTCCGATGTTGACCTGCTGAGT					
		GGCGACCA GGTGATGA AATGA GCTATGA GG TC GATGTTGA CTGCTGAG					
45	-MEXICO	TCGGCGACCAAGGCTGATGATAATGAAGCTATGAGGGGTCCGATGTTGACCTGCTGAGT					
		1570v	1580v	1590v	1600v	1610v	1620v
	-BURMA	CCGCCATTTGTGATATATCTGGGTCTTATGCTGCTCCCTGGCACTGCCCTCCAACCGCTCT					
		C GCCA GAGAT C GG TC TA TCGT TGG C CT CAA C TCT					
50	-MEXICO	CCGCCATTTGTGATATATCTGGGTCTTATGCTGCTCCCTGGCACTGCCCTCCAACCGCTCT					
		1630v	1640v	1650v	1660v	1670v	1680v
	-BURMA	ACCAAGGCTTTGTGCTTTTATGAGTACCTTGGCGGCATTAGCTACAAGGTCA					
		A CA GC CTGGA CT CC GGTGA T GT GCTGCGC G CCG CTG C GC ACAG					
55	-MEXICO	ACCAAGGCTTTGTGCTTTTATGAGTACCTTGGCGGCATTAGCTACAAGGTCA					





		3010v	3020v	3030v	3040v	3050v	3060v
5	-BURMA	TTGTCGTGGTCCGACGCGTGGTTCGTAATGCTGGCGCGGTCGCGGGCTTTGCTGCTT					
	-MEXICO	TTGT GT GT CC AC CG GAG T CG AA GC TGGCG CG CG GGCTTTGC GC T					
		TTGTTGTTGTGCGGACGTCGCGAGCTTCGGAACGCTTGGCGGCGCGCGGGCTTTGCGGCAT					
		3070v	3080v	3090v	3100v	3110v	3120v
	-BURMA	TTACCCCGGACACTGCGCGCGAGATGACCGCAAGGCGCGGGTTGTCATTGATGAGGCTC					
	-MEXICO	T AC CGCGA ACTGC GCG GTCAC CG CG GGGTTGTCATTGATGAGGC C					
10		TGCCTCGCGACGTCGCGCGCGTGTGACTGCGCGCTAGGGTTGTCATTGATGAGGCCC					
		3130v	3140v	3150v	3160v	3170v	3180v
	-BURMA	CATCGCTCGCGCGCTGAGCTGCTGCTGCTGCGACATGCAGCGGGCGCGACCGTCCACCTTC					
	-MEXICO	C TC CTGCGCGG CAC TGGTGT T CA ATGCAGCG GC GC C GT CACCT C					
15		CTTCGCTCGCGCGGACACTGCTGCTTTACATATGCAGCGTGTGCATCTGTGCACCTCC					
		3190v	3200v	3210v	3220v	3230v	3240v
	-BURMA	TTGGCGACCGGAACCGAGATCGCGCGCATCGACTTTGAGCACGCTGGGCTCGTCCCCGCCA					
	-MEXICO	TTGG GACCGGAA CAGATCGG GCGAT GA TTTGAGCAC C GG CT T CC GC A					
20		TTGGGGACCGGAATCGAGATCGCGCGCATAGATTTTGGAGCACACCGGTCTGATTCCAGCAA					
		3250v	3260v	3270v	3280v	3290v	3300v
	-BURMA	TCAGGCGCGACTTAGGCGCGACCTCCTGGTGGCATGTTACCCATCGCTGGCCTGCGGATG					
	-MEXICO	T GGCG GA TT G CGG AC TC TGGTGGCATGT ACCCA CG TG CCTGC GATG					
25		TACGGCGCGGAGTTGGTCCGACTTCATGCTGGCATGTACCCACCGTTGCCCTGCAGATG					
		3310v	3320v	3330v	3340v	3350v	3360v
	-BURMA	TATGCGAGGTCATCCGTGGTGCATACCCCATGATCCAGACCACTAGCCGGGTTCTCCGTT					
	-MEXICO	T TG GAG T TCCGTGGTGC TACCC A ATCCAGAC AC AG GGT CTCGCTT					
30		TCTGTGAGTTAGTCCGTGGTGTACCCCTAAAATCCAGACTACAAGTAAGGTGCTCCGTT					
		3370v	3380v	3390v	3400v	3410v	3420v
	-BURMA	CGTTGTTCTGGGGTGAGCCTGCCGTGCGGCAGAACTAGTGTTCACCCAGGCGGCCAAGC					
	-MEXICO	C T TTCTGGGG GAGCG GC GTCGG CAGAA CTAGTGTTCAC CAGGC GC AAG					
35		CCCTTTTCTGGGGAGAGCCAGCTGTGCGGCAGAAAGCTAGTGTTCACACAGGCTGCTAAGG					
		3430v	3440v	3450v	3460v	3470v	3480v
	-BURMA	CCGCCAACCCCGGCTCAGTGACGGTCCACGAGGCGCAGGGCGCTACCTACACGGAGACCA					
	-MEXICO	CCGC ACCCGGG TC T ACGGTCCA GAGGC CAGGG GC AC T AC AC A					
40		CCGCGCACCCCGGATCTATAACGGTCCATGAGGCCAGGGTGCCACTTTTACCACTACAA					
		3490v	3500v	3510v	3520v	3530v	3540v
	-BURMA	CTATTATTGCCACAGCAGATGCGCGGGGCTTATTGAGTCGTCTCGGGCTCATGCCATTG					
	-MEXICO	CTAT ATTGC AC GCAGATGCGCG GGCCT AT CAGTC TC CGGGCTCA GC AT G					
45		CTATAATTGCAACTGCAGATGCGCGTGGCTCATACAGTCTCCCGGGCTCACGTATAG					
		3550v	3560v	3570v	3580v	3590v	3600v
	-BURMA	TTGCTCTGACGCGCCACACTGAGAAGTGGGTGATCATTGACGCACCCAGGCTGCTTCGCG					
	-MEXICO	TTGCTCT AC G CA ACTGA AA TG GT AT TTGAC C CC GGCCTG T CG G					
50		TTGCTCTGACTAGGCATACTGAAGATGTGTATACCTTGACTCTCCCGGCTGTTGCGTG					
		3610v	3620v	3630v	3640v	3650v	3660v
	-BURMA	AGGTGGGCTATCTCCGATGTAATGCTTAATACTTTTCTCGCTGGTGGCGAAATTGGTC					
	-MEXICO	AGGTGGG ATCTC GATG AT GTTAATAA TT TTCT C GGTGGCGA TTGGTC					
55		AGGTGGGATCTCAGATGCTATGTTAATAAATTTCTCTCTTCGGGTGGCGAGGTTGGTC					



		4210v	4220v	4230v	4240v	4250v	4260v
	-TASHKENT	AGACCATGCGCCATGCTAAAGTGGGCGAGGGCATTTGGCCCTGGAGTAAGACCTTCTGTG					
		AGACCATGCGCCATGCTAAAGTGGGCGAGGGCATTTGGCCCTGGAG AAGACCTTCTG G					
5	-BURMA	AGACCATGCGCCATGCTAAAGTGGGCGAGGGCATTTGGCCCTGGAGCAAGACCTTCTGCG					
		AGAC ATTGC CATGG AAAAT GG CAGGG ATCT CTGGAG AAGAC TT TG G					
	-MEXICO	AGACCAATTGCGCATGCGCAAGTGGGTAGGGTATCTCCGCTGGAGTAAGACGTTTTGTG					
		4270v	4280v	4290v	4300v	4310v	4320v
10	-TASHKENT	CCCTTTTGGGCGCCCTGGTTCCGTGCTATTGAGAAGGCTATTCTGGCCCTGCTCCCTCAGG					
		CCCT TT GGGCG TGGTTCCG GCTATTGAGAAGGCTATTCTGGCCCTGCTCCCTCAGG					
	-BURMA	CCCTTTTGGGCGCTGGTTCCGCGCTATTGAGAAGGCTATTCTGGCCCTGCTCCCTCAGG					
		CCCT TTTGGCGG TGGTTCCG GC ATTGAGAAGGCTATTCT CCCT T CC CA G					
	-MEXICO	CCCTGTTTGGGCGCTGGTTCCGTGCGATTGAGAAGGCTATTCTATCCCTTTTACCACAAG					
		4330v	4340v	4350v	4360v	4370v	4380v
15	-TASHKENT	GTGTGTTTTATGGGGATGCTTTTGATGACACCGTCTTCTCGGCGCGTGTGGCCGAGCAA					
		GTGTGTTTTA GG GATGCGTTTGATGACACCGTCTTCTCGGCG GTGGCCGAGCAA					
	-BURMA	GTGTGTTTTACGGTGATGCTTTTGATGACACCGTCTTCTCGGCGCGTGTGGCCGAGCAA					
		GTGTGTT TACGG GATGC T TGA GAC C GT TTCTC GC GC GTGGC G GC A					
20	-MEXICO	CTGTGTTCTACGGGATGCTTTATGACGACTCAGTATTCTCTGCTGCCGTGGCTGGCGCCA					
		4390v	4400v	4410v	4420v	4430v	4440v
25	-TASHKENT	AGGCGTCCATGGTGTGTTTGAAGTGAAGTCTTCTGAGTTTGACTCCACCCAGAATAATTTTT					
		AGGC TCCATGGTGTGTTTGAAGTGAAGTCTTCTGAGTTTGACTCCACCCAGAATAA TTTT					
	-BURMA	AGGCGTCCATGGTGTGTTTGAAGTGAAGTCTTCTGAGTTTGACTCCACCCAGAATAACTTTT					
		CCATGGTGTGTTGA AATGA TTTTCTGAGTTTGACTC AC CAGAATAACTTTT					
	-MEXICO	GCCATGCCATGGTGTGTTTGAAGTGAAGTCTTCTGAGTTTGACTCGACTCAGAATAACTTTT					
		4450v	4460v	4470v	4480v	4490v	4500v
30	-TASHKENT	CCCTGGGCTAGAGTGTGCTATTATGGAGAAGTGTGGGATGCCGAAGTGGCTCATCCGCT					
		C CTGGG CTAGAGTGTGCTATTATGGAG AGTGTGGGATGCCG AGTGGCTCATCCGC					
	-BURMA	CTCTGGGCTAGAGTGTGCTATTATGGAGGAGTGTGGGATGCCGAGTGGCTCATCCGCC					
		C CT GGTCT GAGTG GC ATTATGGA GAGTGTGG ATGCC CAGTGGCT TC G					
35	-MEXICO	CCCTAGGTCTTGAAGTGGCGCATTTATGGAAGAGTGTGGTATGCCCAAGTGGCTGTCCAGGT					
		4510v	4520v	4530v	4540v	4550v	4560v
40	-TASHKENT	TGTACCACCTTATAAGTCTCGCGTGGATCCTGCAGGCCCCGAAGGAGTCCCTGCGAGGGT					
		TGTA CACCTTATAAGTCTCGCGTGGATC TGCAGGCCCCGAAGGAGTC CTGCGAGGGT					
	-BURMA	TGTATCACCTTATAAGTCTCGCGTGGATCCTGCAGGCCCCGAAGGAGTCTCTGCGAGGGT					
		TGTA CA T GGTG GCGTGGATC TGCAGGCCCC AA GAGTCT TG GAGGGT					
	-MEXICO	TGTACCATGCCGTCCGCTCGCGTGGATCCTGCAGGCCCCAAAGAGTCTTTGAGAGGGT					
		4570v	4580v	4590v	4600v	4610v	4620v
45	-TASHKENT	GTGGAAGAAACACTCCGGTGAAGCCCGGCACTCTTCTATGGAATACTGTCTGGAACATGG					
		TTGGAAGAAACACTCCGGTGAAGCCCGGCACTCTTCTATGGAATACTGTCTGGAATGG					
	-BURMA	TTTGAAGAAACACTCCGGTGAAGCCCGGCACTCTTCTATGGAATACTGTCTGGAATATGG					
		T TGAAGAA CA TC GGTGAGCC GGCA T CT TGAATAC GT TGAATGG					
	-MEXICO	TGTGGAAGAGCATTTCTGTTGAGCCGGGCACTTGTCTGGAATACGGTGTGGAACATGG					
		4630v	4640v	4650v	4660v	4670v	4680v
50	-TASHKENT	CCGTTATACCCATTTTACGATTTCCGCGATTTGCAGGTGGCTGCCTTTAAAGGTGATG					
		CCGTTAT ACCCA TGTGA GA TTCCGCGATTT AGGTGGCTGCCTTTAAAGGTGATG					
	-BURMA	CCGTTATACCCACTGTTATGACTTCCGCGATTTTCAAGGTGGCTGCCTTTAAAGGTGATG					
		C T ATT CCGA TG TATGA TTCCG GA T CAGGT GC GCCTT AA GG GA G					
55	-MEXICO	CAATCATTTGCCCATTTGCTATGAGTTCCGGGACCTCCAGGTTGCCGCTTCAAGGGCGACG					

4690v 4700 4710v 4720v 4730v 4740v  
 5 -TASHKENT ATTGATAGTGTCTTTGCAATGAGTACCGTACAGTCCAGGGGCTGCTGTCTGATTGCTG  
 ATTGATAGTGTCTTTGCAATGAGTACCGTACAGTCCAGG GCTGCTGTCTGAT GC G  
 -BURMA ATTGATAGTGTCTTTGCAATGAGTACCGTACAGTCCAGGAGCTGCTGTCTGATCGCCG  
 A TCG T GT CT TG AGTA TA CG CAGAG CCAGG GC G T CT AT GC G  
 -MEXICO ACTCGGTGTCTCTGTATGAAATCCGCTACAGTCCAGGGGCGCGGTTCTGCTTATAGCAG

4750v 4760v 4770v 4780v 4790v 4800v  
 10 -TASHKENT GCTGTGGCTTAAGGTGAAAGTGGGTTTCCGTCGATTTGTTGTATGCAGGTGTTGTGG  
 GCTGTGGCTT AAG TGAAGT G TTTCCG CCGAT GGTGTTGTATGCAGGTGTTGTGG  
 -BURMA GCTGTGGCTTGAAGTGAAGTGAATTTCCGTCGATTTGTTGTATGCAGGTGTTGTGG  
 GCTGTGG TGAAGTGAAGT GA TTTCCG CCGAT GG TGTATGC GG GTTGT G  
 -MEXICO GCTGTGGTTGAAGTGAAGGTGACTTTCCGTCGATTTGTTGTATGCAGGTGTTGTGG

4810v 4820v 4830v 4840v 4850v 4860v  
 20 -TASHKENT TGACCCCGGCTTGGCGCGCTTCCGACGTGCTGCGCTTGTCCGGCCGGCTTACTGAGA  
 TG CCCCCGGCTTGGCGCGCTTCCCGA GT GTGCGCTTG CCGGCCGGCTTAC GAGA  
 -BURMA TGGCCCGCGCTTGGCGCGCTTCCGATGTTGTGCGCTTCCGGCCGGCTTACCGAGA  
 T GCGCC GG CT GG GC CT CC GATGT GT CG TTTCCCGG CCGCTT C GAGA  
 -MEXICO TCGCCCCGGGCTTGGCGCGCTTCCCGATGTTGTGTTGATTGCGCCGACGGCTTTCGGAGA

4870v 4880v 4890v 4900v 4910v 4920v  
 25 -TASHKENT AGAATTGGGGCCCTGGCCCTGAGCGGGGAGCAGCTCCGCCCTTGCTGT  
 AGAATTGGGGCCCTGGCCCTGAGCGGGGAGCAGCTCCGCCCT GCTGT  
 -BURMA AGAATTGGGGCCCTGGCCCTGAGCGGGGAGCAGCTCCGCCCTGCTGTGTTAGTGATTTCC  
 AGAA TGGGG CCTG CC GAGCGGGC GAGCAGCTCCGCCCTGC GT GATTTCC  
 -MEXICO AGAATTGGGGCCCTGATCCGAGCGGGGAGCAGCTCCGCCCTGCGGTGACAGGATTTCC

4930v 4940v 4950v 4960v 4970v 4980v  
 30 -BURMA TCCGCAAGCTCACGAATGTAGCTCAGATGTGTGTTGGATGTTGTTTCCCGTGTGTTATGGGG  
 TCCG A G T ACGAATGT GC CAGAT TGTGT GA GT GT TC G GTTTA GGGG  
 -MEXICO TCCGTAGGTTAACGAATGTGGCCCGAGATTTGTGTTGAGGTGGTGTCTAGAGTTTACGGGG

4990v 5000v 5010v 5020v 5030v 5040v  
 35 -BURMA TTTCCCTGGACTGTTTATAACCTGATTGGCATGCTACAGGCTGTTGCTGATGGCAAGG  
 TTTCCCC GG CT GTTATAACCTGAT GGCATGCT CAG CT TTG TGATGG AAGG  
 -MEXICO TTTCCCGGGTCTGTTTATAACCTGATAGGCATGCTCCAGACTATTGGTGATGGTAAGG

5050v 5060v 5070v 5080v 5090v 5100v  
 40 -BURMA CACATTTCACTGAGTCAGTAAAACAGTGCCTGACTTGACAAATTCAATCTTGTGTGCGG  
 C CATTT AC GAGTC GT AA CC T CT GAC T ACA A TCAAT TG CGG  
 -MEXICO CGCATTTTACAGAGTCTGTTAAGCCTATACTTGACCTTACACACTCAATTATGCACCGGT

5110v 5120 5130v 5140v 5150v 5160v  
 45 -BURMA TGGAAATGAATAACATGTCTTTTGTGCTGCGCCCATGGGTTGCGGACCATGCGCCCTCGGCCT  
 GAATGAATAACATGT TTTGCTGCGCCCATGGGTTGCG ACCATGCGCCCT GGCCT  
 -MEXICO CTGAATGAATAACATGTGTTTGTGCTGCGCCCATGGGTTGCGGACCATGCGCCCTAGGCCT

5170v 5180v 5190v 5200v 5210v 5220v  
 50 -BURMA ATTTTGTGCTGCTCTCTATGTTTGTGCTATGCTGCGCGCGCCACCGCCCGGTGACGCC  
 TTTTG TG TG TCGTC TGTTC TGCCTATG TCGCGCGCCACCG CCGGTGACGCC  
 -MEXICO CTTTGTGCTGTTGTTCTCTCTATGTTTGTGCTATGTTGCGCGCGCCACCGCCCGGTGACGCC

		5230v	5240v	5250v	5260v	5270v	5280v
	-BURMA	TCTGGCGCGCGTGGTGGGCGCGCGAGCGCGGTTCCGGCGGTGGTTTCTGGGGTGACCGG					
		TCTGGCGCGCGTGGTGGGCGCGCGAGCGCGGTTCCGGCGGTGGTTTCTGGGGTGACCGG					
	-MEXICO	TCTGGCGCGCGTGGTGGGCGCGCGAGCGCGGTTCCGGCGGTGGTTTCTGGGGTGACCGG					
5		5290v	5300v	5310v	5320v	5330v	5340v
	-BURMA	GTTGATTCTCAGCGCTTGGAAATCCCGTATATTCATCCAACCAACCCCTTCGCCCGCGAT					
		GTTGATTCTCAGCGCTTGGAAATCCCGTATATTCATCCAACCAACCCCTTCGCCCGCGAT					
	-MEXICO	GTTGATTCTCAGCGCTTGGAAATCCCGTATATTCATCCAACCAACCCCTTCGCCCGCGAT					
10		5350v	5360v	5370v	5380v	5390v	5400v
	-BURMA	GTCACCGCTGGCGCGCGCGTGGACCTCGTGTTCGGCAACCGCGCCGACCACTCGGCTCC					
		GT CGCGTGGG CGCGG CTGGACCTCG GTTCGGCAACCG CGCG CCACT GGCTCC					
	-MEXICO	GTTGGCGCTGGCTCGCGGCTGGACCTCGTGTTCGGCAACCGCGCCGACCACTCGGCTCC					
15		5410v	5420v	5430v	5440v	5450v	5460v
	-BURMA	GCTTGGCGTGACCAAGCGCGCGCGCGGTTGCCCTCACGTCGTAGACCTACCAAGCT					
		CTTGGCG GA CAGGCGCGCGCGCG CG TGCGTC CGTCG GACCT CCACAGC					
	-MEXICO	ACTTGGCGAGATCAGGCGCGCGCGCGGCTCGCTCGCGGACCTGCCACAGCC					
20		5470v	5480v	5490v	5500v	5510v	5520v
	-BURMA	GGGGCGCGCGCGCTAAACCGCGGTCGCTCCGGCCCATGACACCCCGCCAGTGCCTGATGTC					
		GGGGC GCG CGCT AC GC GT GC CC GCCCATGACACC C CC GT CC GA GT					
	-MEXICO	GGGGCTGGCGCGCTGACGGCTGTGGCGCCTGCCCATGACACCTCACCCGTCGCGACGTT					
25		5530v	5540v	5550v	5560v	5570v	5580v
	-BURMA	GACTCCCGCGCGCGCATCTTGGCGCGGAGTATAACCTATCAACATCTCCCTTACCTCT					
		GA TC CGCGG GC AT T CGCGG CAGTATAA T TC AC TC CCCCT AC TC					
	-MEXICO	GATTCTCGCGGTGCAATTCTACGCGCGGAGTATAATTTGTCTACTTACCCCTGACATCC					
30		5590v	5600v	5610v	5620v	5630v	5640v
	-BURMA	TCCGTGGCCACCGGCACTAACCTGGTTCTTTATGCCGCCCCCTTTAGTCCGCTTTTACCC					
		TC GTGGCC C GGCACTAA T GT CT TATGC GCCCG CTTA TCCGC T T CC					
	-MEXICO	TCTGTGGCCTCTGGCACTAATTTAGTCTGTATGCAGCCCCCTTAATCCGCTCTGCCG					
35		5650v	5660v	5670v	5680v	5690v	5700v
	-BURMA	CTTCAGGACGGCACCAATACCATATAATGGCCACGGAAGCTTCTAATTATGCCAGTAC					
		CT CAGGACGG AC AATAC CA AT ATGGCCAC GA GC TC AATTATGC CAGTAC					
	-MEXICO	CTGCAGGACGGTACTAATACTCACATTATGGCCACAGAGGCTCCAATTATGCACAGTAC					
40		5710v	5720v	5730v	5740v	5750v	5760v
	-BURMA	CGGGTTGCCCGTGCCCAATCCGTTACCGCCCGCTGGTCCCCAATGCTGTGGCGGTTAC					
		CGGGTTGCCCG GC AC ATCCGTTACCG CC CT GT CC AATGC GT GG GG TA					
	-MEXICO	CGGGTTGCCCGCGCTACTATCCGTTACCGCCCGCTAGTGCCTAATGCAGTTGGAGGCTAT					
45		5770v	5780v	5790v	5800v	5810v	5820v
	-BURMA	GCCATCTCCATCTCTTCTGGCCACAGACCAACCAACCCCGACGTCGTTGATATGAAT					
		GC AT TCCAT TC TTCTGGCC CA AC ACCAC ACCCC AC TC GTTGA ATGAAT					
	-MEXICO	GCTATATCCATTTCTTCTGGCCCAACCAACCAACCCCTACATCTGTTGACATGAAT					
50		5830v	5840v	5850v	5860v	5870v	5880v

		5830v	5840v	5850v	5860v	5870v	5880v
	-BURMA	TCAATAACCTGGACGGATGTTGGTATTTTGTCCAGCCCGGCATAGCCTCTGAGCTTGTG					
	-MEXICO	TC AT AC TC AC GATGT G ATT T GT CA CC GGCATAGC TCTGA T GT					
5							
		5890v	5900v	5910v	5920v	5930v	5940v
	-BURMA	ATCCCAAGTGAGCGCCTACACTATCGTAACCAAGGCTGGCGCTCCGTCGAGACCTCTGGG					
	-MEXICO	ATCCCAAG GAGCGCCT CACTA CG AA CAAGG TGGCGCTC GT GAGAC TCTGG					
10							
		5950v	5960v	5970v	5980v	5990v	6000v
	-BURMA	GTGGGTGAGGAGGAGGCTACCTCTGCTCTTTATGCTTTGCATACATGGCTCACTCGTA					
	-MEXICO	GT GCTGAGGAGGA GC ACCTC GGTCTGT ATG T TGCATACATGGCTC C GT					
15							
		6010v	6020v	6030v	6040v	6050v	6060v
	-BURMA	AATTCTTATACTAATACACCTTATACCGGTGGCGCTGTTGGACTTTGCCCTTGAG					
	-MEXICO	AA TCTATAC AATAC CC TATACCGGTGGCGT GG T TGGACTTTGCC T GAG					
20							
		6070v	6080v	6090v	6100v	6110v	6120v
	-BURMA	CTTGAGTTTTCGCAACCTTACCCCGGTACACCAATACGCGGGTCTCCCGTTATTCAGC					
	-MEXICO	CTTGAGTTTTCGCAA CT ACC CC GTAACACCAATAC CG GT TCCCGTTA TCCAGC					
25							
		6130v	6140v	6150v	6160v	6170v	6180v
	-BURMA	ACTGCTCGCCACCGCCTTCGTGCGGTGCGGACGGGACTGCCGAGCTCACCACCACGGCT					
	-MEXICO	ACTGCTCG CAC C CG G G GACGGGACTGC GAGCT ACCAC AC GC					
30							
		6190v	6200v	6210v	6220v	6230v	6240v
	-BURMA	GCTACCCGCTTTATGAAGGACCTCTATTTTACTAGTACTAATGGTGTGCGTGAGATCGGC					
	-MEXICO	GC ACC G TT ATGAA GA CTC A TTTAC G TAATGG GT GGTGA TCGGC					
35							
		6250v	6260v	6270v	6280v	6290v	6300v
	-BURMA	CGCGGGATAGCCCTCACCTGTTCAACCTTGCTGACACTCTGCTTGGCGGCTGCCGACA					
40							
	-MEXICO	CGCGGGATAGC CT AC T T AACCTTGCTGACAC CT CT GGCGG CT CCGACA					
		6310v	6320v	6330v	6340v	6350v	6360v
	-BURMA	GAATTGATTTGTCGGCTGGTGGCGAGCTGTTCTACTCCCGTCCCGTTGTCTCAGCCAAT					
45							
	-MEXICO	GAATT ATTTGTCGGCTGG GG CA CTGTT TA TCCCG CC GTTGTCTCAGCCAAT					
		6370v	6380v	6390v	6400v	6410v	6420v
	-BURMA	GGCGAGCCGACTGTTAAGTTGTATACATCTGTAGAGAATGCTCAGCAGGATAAGGGTATT					
50							
	-MEXICO	GGCGAGCC AC GT AAG T TATACATC GT GAGAATGCTCAGCAGGATAAGGGT TT					
		6430v	6440v	6450v	6460v	6470v	6480v
	-BURMA	GCAATCCCGCATGACATTGACCTCGGAGAATCTCGTGTGGTTATTTCAGGATTATGATAAC					
55							
	-MEXICO	GC ATCC CA GA AT GA CT GG GA TC CGTGTGGT ATTCAGGATTATGA AAC					

		6490v	6500v	6510v	6520v	6530v	6540v
	-BURMA	CAACATGAACAAATCGGCGGAGTGGTCTCTGAGCGCCATCGCGCCCTTTCTCTGTCCTT					
5	-MEXICO	CA CATGA CA GATGCGT AD CG TC CG GC CCATC CG CCTTT TCTGT CT					
		6550v	6560v	6570v	6580v	6590v	6600v
	-BURMA	CGAGCTAATGATGTGCTTGGCTCTCTGAGCGCTGCGGAGTATGACCAGTCCACTTAT					
10	-MEXICO	CGAGC AATGATGT GTTGGCT TC CTGAC GC GCGAGTATGACCAGTCCACTTA					
		6610v	6620v	6630v	6640v	6650v	6660v
	-BURMA	GGCTCTTGGACTGGCGGAGTTTATGTTTCTGACTCTGTGACCTTGGTTAATGTTGCGACC					
15	-MEXICO	GG TC TC ACTGGCGG GTTTAT T TC GAC GTGAC TTGGT AATGTTGCGAC					
		6670v	6680v	6690v	6700v	6710v	6720v
	-BURMA	GGCGCGCAGGCGGTTGGCGGTCGCTCGATTGGACCAAGGTACACTTGACGGTCGCCCC					
20	-MEXICO	GGCGCGCAGGCGGTGGCGGTCGCTCGATTGGACCAAGGTACACTTGACGGTCGCCCC					
		6730v	6740v	6750v	6760v	6770v	6780v
	-BURMA	CTCTCCACCATCCAGCAGTACTCGAAGACCTTCTTTGTCTGCCGCTCCGCGTAAGCTC					
25	-MEXICO	CTC C AC T AGCA TA TC AAGAC TTCTTTGT CT CC CT CG GG AAGCTC					
		6790v	6800v	6810v	6820v	6830v	6840v
	-BURMA	TCTTTCTGGGAGGCGAGGCACAACTAAAGCCGGGTACCTTATAATTATAACCACTGCT					
30	-MEXICO	TC TT TGGGAGGC GGCACAC AAAGC GG TA CCTTATAATTATAA AC ACTGCT					
		6850v	6860v	6870v	6880v	6890v	6900v
	-BURMA	AGCGACCAACTGCTTGTGAGGAATGCCGCCGGGCACCGGTCGCTATTTCCACTTACACC					
35	-MEXICO	AG GACCA T CT T GA AATGC GCCGG CA CGGGTCGC ATTTT AC TA ACC					
		6910v	6920v	6930v	6940v	6950v	6960v
	-BURMA	ACTAGCCTGGGTGCTGGTCCCGTCTCCATTTCTGCGGTTGCCGTTTTAGCCCCCACTCT					
40	-MEXICO	AC AG CT GG GC GGTCC GTC CCATTTCTGCGG GC GTTTT GC CC C CTC					
		6970v	6980v	6990v	7000v	7010v	7020v
	-BURMA	GCGCTAGCATTGCTTGAGGATACCTTGGACTACCTGCCCGCGCCATACTTTTGATGAT					
45	-MEXICO	GC CT GC TGCT GAGGATAC TT GA TA CC G CG GC CA AC TTTGATGA					
		7030v	7040v	7050v	7060v	7070v	7080v
	-BURMA	TTCTGCCGAGAGTGGCGCGGCTTGGCTTCAGGGGTCGCTTTCCAGTCTACTGTCGCT					
50	-MEXICO	TTCTGCCG GA TGCGG C T GGCCT CAGGG TG GCTTTCCAGTC ACTGTCGCT					
		7090v	7100v	7110v	7120v	7130v	7140v
	-BURMA	GAGCTTCAGCGCCTTAAGATGAGGTGGGTAAAACTCGGGAGTTGTAGTTTATTTGCTTG					
55	-MEXICO	GAGCT CAGCGCCTTAA T AAGGTGGGTAAAACTCGGGAGTTGTAGTTTATTTGCTG					

```

              7165v      7167v                7170v      7180v      7190v
-BURMA    TGGCCGCCGCTTCTTTCTGTTC-----TTATTCTCATTTCTGC GTTCCGCGCTCCC
          TGGCC GCT CTT       TGC             TTATTC   TTCT  GT CCGCGCTCCC
-MEXICO    TGGCCACGCTACTTATACTGCTGATTTCTTTATTTCTTTTTCTCGGTCCC GCGCTCCC
           ✓ 7195
-BURMA     TGA
           TGA
-MEXICO     TGA

```

10

A number of open reading frames, which are potential coding regions, have been found within the DNA sequences set forth above. As has already been noted, consensus residues for the RNA-directed RNA polymerase (RDRP) were identified in the HEV (Burma) strain clone ET1.1. Once a contiguous overlapping set of clones was accumulated, it became clear that the nonstructural elements containing the RDRP as well as what were identified as consensus residues for the helicase domain were located in the first large open reading frame (ORF1). ORF1 covers the 5' half of the genome and begins at the first encoded met, after the 27th bp of the apparent non-coding sequence, and then extends 5079 bp before reaching a termination codon. Beginning 37 bp downstream from the ORF1 stop codon in the plus 1 frame is the second major opening reading frame (ORF2) extending 1980 bp and terminating 68 bp upstream from the point of poly A addition. The third forward ORF (in the plus 2 frame) is also utilized by HEV. ORF3 is only 370 bp in length and would not have been predicted to be utilized by the virus were it not for the identification of the immunoreactive cDNA clone 406.4-2 from the Mexico SISPA cDNA library (see below for detailed discussion). This epitope confirmed the utilization of ORF3 by the virus, although the means by which this ORF is expressed has not yet been fully elucidated. If we assume that the first met is utilized, ORF3 overlaps ORF1 by 1 bp at its 5' end and ORF2 by 328 bp at its 3' end. ORF2 contains the broadly reactive 406.3-2 epitope and also

20

25

30

35

40

a signal sequence at its extreme 5' end. The first half of this ORF2 also has a high pI value (>10) similar to that seen with other virus capsid proteins. These data suggest that the ORF2 might be the predominant structural gene of HEV.

5 The existence of subgenomic transcripts prompted a set of experiments to determine whether these RNAs were produced by splicing from the 5' end of the genome. An analysis using subgenomic probes from throughout the genome, including the extreme 5' end, did not provide evidence for a spliced transcript. However, it was discovered that a region of the genome displayed a high degree of homology with a 21 bp segment identified in Sindbis as a probably internal initiation site for RNA transcription used in the production of its subgenomic messages. Sixteen of 21 (76%) of the nucleotides are identical.

Two cDNA clones which encode an epitope of HEV that is recognized by sera collected from different ET-NANB outbreaks (i.e., a universally recognized epitope) have been isolated and characterized. One of the clones immunoreacted with 8 human sera from different infected individuals and the other clone immunoreacted with 7 of the human sera tested. Both clones immunoreacted specifically with cyno sera from infected animals and exhibited no immunologic response to sera from uninfected animals. The sequences of the cDNAs in these recombinant phages, designated 406.3-2 and 406.4-2 have been determined. The HEV open reading frames are shown to encode epitopes specifically recognized by sera from patients with HEV infections. The cDNA sequences and the polypeptides that they encode are set forth below.

Epitopes derived from Mexican strain of HEV:

35 406.4-2 sequence (nucleotide sequence has SEQ ID NO.13; amino acid sequence has SEQ ID NO.14):

SEQ ID NO. 13:

5	C GCC AAC CAG CCC GGC GAG TTG GTT TTA GTT GGC GAG ATC AGG CCC Ala Asn Gln Pro Gly His Leu Ala Pro Leu Gly Glu Ile Arg Pro 1 5 10 15	46
10	AGC GGC CCT CCG CTG COT CCC GTC GCG GAG CTG CCA CAG CCG GGG CTG Ser Ala Pro Pro Leu Pro Pro Val Ala Asp Leu Pro Gln Pro Gly Leu 20 25 30	94
15	CGG CGC TGA CCGCTGTGGC GCCTGCCCAT GACACCTCAC CCGTCCCGGA Arg Arg .	143
20	CGTTGATTCT CGCGGTGCAA TTCTAGCCCG CAGTATAAT TTGTCTACTT CACCCCTGAC ATCCTCTGTG GCCTCTGGCA CTAATTTAGT COTSTATGCA GCCGCCCTTA ATCCGCCTCT GCCGCTGCAG GACGGTACTA ATACTACAT TATGGCCACA GAGGCTTCCA ATTATGCACA GTACCGGGTT GCCCGCGCTA CTATCGGTA CCGGCCCTA GTGCCTAATG CAGTTGGAGG CTATGCTATA TCCATTTCTT TCTGGCCTCA AACACCCACA ACCCCTACAT CTGTTGACAT GAATTC	203 263 323 383 443 449

SEQ ID NO. 14:

30	Ala Asn Gln Pro Gly His Leu Ala Pro Leu Gly Glu Ile Arg Pro Ser 1 5 10 15
35	Ala Pro Pro Leu Pro Pro Val Ala Asp Leu Pro Gln Pro Gly Leu Arg 20 25 30
	Arg .

406.3-2 sequence (nucleotide sequence has SEQ ID NO.15; amino acid sequence has SEQ ID NO.16):

SEQ ID NO. 15:

40	GGAT ACT TTT GAT TAT CCG GGG CCG GCG CAC ACA TTT GAT GAC TTC TGC Thr Phe Asp Tyr Pro Gly Arg Ala His Thr Phe Asp Asp Phe Cys 1 5 10 15	49
45	CCT GAA TGC CGC GCT TTA GGC CTC CAG GGT TGT GCT TTC CAG TCA ACT Pro Glu Cys Arg Ala Leu Gly Leu Gln Gly Cys Ala Phe Gln Ser Thr 20 25 30	97
50	GTC GCT GAG CTC CAG CGC CTT AAA GTT AAG GTT Val Ala Glu Leu Gln Arg Leu Lys Val Lys Val 35 40	130



the complementary DNA sequence. Additionally, open reading frames encoding peptides are present, and expressible peptides are disclosed by the nucleotide sequences without setting forth the amino acid sequences explicitly, in the same manner as if the amino acid sequences were explicitly set forth as in the ET1.1 sequence or other sequences above.

## DETAILED DESCRIPTION OF THE INVENTION

### 10 I. Definitions

The terms defined below have the following meaning herein:

15 1. "Enterically transmitted non-A/non-B hepatitis viral agent, ET-NANB, or HEV" means a virus, virus type, or virus class which (i) causes water-borne, infectious hepatitis, (ii) is transmissible in cynomolgus monkeys, (iii) is serologically distinct from hepatitis A virus (HAV), hepatitis B virus (HBV), hepatitis C virus (HCV), and hepatitis D virus, and  
20 (iv) includes a genomic region which is homologous to the 1.33 kb cDNA insert in plasmid pTZKF1(ET1.1) carried in E. coli strain BB4 identified by ATCC deposit number 67717.

25 2. Two nucleic acid fragments are "homologous" if they are capable of hybridizing to one another under hybridization conditions described in Maniatis et al., op. cit., pp. 320-323. However, using the following wash conditions: 2 x SCC, 0.1% SDS, room temperature twice, 30 minutes each; then 2 x SCC, 0.1%  
30 SDS, 50°C once, 30 minutes; then 2 x SCC, room temperature twice, 10 minutes each, homologous sequences can be identified that contain at most about 25-30% basepair mismatches. More preferably, homologous nucleic acid strands contain 15-25%  
35 basepair mismatches, even more preferably 5-15% basepair mismatches. These degrees of homology can be selected by using more stringent wash conditions for identification of clones from gene libraries (or

other sources of genetic material), as is well known in the art.

3. Two amino acid sequences or two nucleotide sequences (in an alternative definition for homology between two nucleotide sequences) are considered homologous (as this term is preferably used in this specification) if they have an alignment score of  $>5$  (in standard deviation units) using the program ALIGN with the mutation gap matrix and a gap penalty of 6 or greater. See Dayhoff, M.O., in Atlas of Protein Sequence and Structure (1972) Vol. 5, National Biomedical Research Foundation, pp. 101-110, and Supplement 2 to this volume, pp. 1-10. The two sequences (or parts thereof, preferably at least 30 amino acids in length) are more preferably homologous if their amino acids are greater than or equal to 50% identical when optimally aligned using the ALIGN program mentioned above.

4. A DNA fragment is "derived from" an ET-NANB viral agent if it has the same or substantially the same basepair sequence as a region of the viral agent genome.

5. A protein is "derived from" an ET-NANB viral agent if it is encoded by an open reading frame of a DNA or RNA fragment derived from an ET-NANB viral agent.

## II. Obtaining Cloned ET-NANB Fragments

According to one aspect of the invention, it has been found that a virus-specific DNA clone can be produced by (a) isolating RNA from the bile of a cynomolgus monkey having a known ET-NANB infection, (b) cloning the cDNA fragments to form a fragment library, and (c) screening the library by differential hybridization to radiolabeled cDNAs from infected and non-infected bile sources.

### A. cDNA Fragment Mixture

ET-NANB infection in cynomolgus monkeys is initiated by inoculating the animals intravenously with a 10% w/v suspension from human case stools positive for 27-34 nm ET-NANB particles (mean diameter 32 nm). An infected animal is monitored for elevated levels of alanine aminotransferase, indicating hepatitis infection. ET-NANB infection is confirmed by immunospecific binding of seropositive antibodies to virus-like particles (VLPs), according to published methods (Gravelle). Briefly, a stool (or bile) specimen taken from the infected animal 3-4 weeks after infection is diluted 1:10 with phosphate-buffered saline, and the 10% suspension is clarified by low-speed centrifugation and filtration successively through 1.2 and 0.45 micron filters. The material may be further purified by pelleting through a 30% sucrose cushion (Bradley). The resulting preparation of VLPs is mixed with diluted serum from human patients with known ET-NANB infection. After incubation overnight, the mixture is centrifuged overnight to pellet immune aggregates, and these are stained and examined by electron microscopy for antibody binding to the VLPs.

ET-NANB infection can also be confirmed by seroconversion to VLP-positive serum. Here the serum of the infected animal is mixed as above with 27-34 nm VLPs isolated from the stool specimens of infected human cases and examined by immune electron microscopy for antibody binding to the VLPs.

Bile can be collected from ET-NANB positive animals by either cannulating the bile duct and collecting the bile fluid or by draining the bile duct during necropsy. Total RNA is extracted from the bile by hot phenol extraction, as outlined in Example 1A. The RNA fragments are used to synthesize corresponding duplex cDNA fragments by random priming, also as referenced in Example 1A. The cDNA fragments may be fractionated by gel electrophoresis or density

gradient centrifugation to obtain a desired size class of fragments, e.g., 500-4,000 basepair fragments.

Although alternative sources of viral material, such as VLPs obtained from stool samples (as described in Example 4), may be used for producing a CDNA fraction, the bile source is preferred. According to one aspect of the invention, it has been found that bile from ET-NANB-infected monkeys shows a greater number of intact viral particles than material obtained from stool samples, as evidenced by immune electron microscopy. Bile obtained from an ET-NANB infected human or cynomolgus macaque, for use as a source of ET-NANB viral protein or genomic material, or intact virus, forms part of the present invention.

15

#### B. cDNA Library and Screening

The cDNA fragments from above are cloned into a suitable cloning vector to form a cDNA library. This may be done by equipping blunt-ended fragments with a suitable end linker, such as an EcoRI sequence, and inserting the fragments into a suitable insertion site of a cloning vector, such as at a unique EcoRI site. After initial cloning, the library may be re-cloned, if desired, to increase the percentage of vectors containing a fragment insert. The library construction described in Example 1B is illustrative. Here cDNA fragments were blunt-ended, equipped with EcoRI ends, and inserted into the EcoRI site of the lambda phage vector gt10. The library phage, which showed less than 5% fragment inserts, was isolated, and the fragment inserts re-cloned into the lambda gt10 vector, yielding more than 95% insert-containing phage.

The cDNA library is screened for sequences specific for ET-NANB by differential hybridization to cDNA probes derived from infected and non-infected sources. cDNA fragments from infected and non-infected source bile or stool viral isolates can be prepared as above. Radiolabeling the fragments is by random

35

labeling, nick translation, or end labeling,  
according to conventional methods (Maniatis, p. 109).  
The cDNA library from above is screened by transfer to  
duplicate nitrocellulose filters, and hybridization  
5 with both infected-source and non-infected-source  
(control) radiolabeled probes, as detailed in Example  
2. In order to recover sequences that hybridize at the  
preferred outer limit of 25-30% basepair mismatches,  
clones can be selected if they hybridize under the  
10 conditions described in Maniatis et al., op. cit., pp.  
320-323, but using the following wash conditions: 2 x  
SCC, 0.1% SDS, room temperature - twice, 30 minutes  
each; then 2 x SCC, 0.1% SDS, 50°C - once, 30 minutes;  
then 2 x SCC, room temperature - twice, 10 minutes  
15 each. These conditions allowed identification of the  
Mexican isolate discussed above using the ET1.1  
sequence as a probe. Plaques which show selective  
hybridization to the infected-source probes are  
preferably re-plated at low plating density and re-  
20 screened as above, to isolate single clones which are  
specific for ET-NANB sequences. As indicated in  
Example 2, sixteen clones which hybridized  
specifically with infected-source probes were  
identified by these procedures. One of the clones,  
25 designated lambda gt101.1, contained a 1.33 kilobase  
fragment insert.

### C. ET-NANB Sequences

The basepair sequence of cloned regions of the  
30 ET-NANB fragments from Part B are determined by  
standard sequencing methods. In one illustrative  
method, described in Example 3, the fragment insert  
from the selected cloning vector is excised, isolated  
by gel electrophoresis, and inserted into a cloning  
35 vector whose basepair sequence on either side of the  
insertion site is known. The particular vector  
employed in Example 3 is a pTZKF1 vector shown at the  
left in Figure 1. The ET-NANB fragment from the gt10-

1.1 phage was inserted at the unique EcoRI site of the pTZKF1 plasmid. Recombinants carrying the desired insert were identified by hybridization with the isolated 1.33 kilobase fragment, as described in Example 3. One selected plasmid, identified as pTZKF1 (ET1.1), gave the expected 1.33 kb fragment after vector digestion with EcoRI. E. coli strain BB4 infected with the pTZKF1(ET1.1) plasmid has been deposited with the American Type Culture Collection, Rockville, MD, and is identified by ATCC deposit number 67717.

The pTZKF1(ET1.1) plasmid is illustrated at the bottom in Figure 1. The fragment insert has 5' and 3' end regions denoted at A and C, respectively, and an intermediate region, denoted at B. The sequences in these regions were determined by standard dideoxy sequencing and were set forth in an earlier application in this series. The three short sequences (A, B, and C) are from the same insert strand. As will be seen in Example 3, the B-region sequence was actually determined from the opposite strand, so that the B region sequence shown above represents the complement of the sequence in the sequenced strand. The base numbers of the partial sequences are approximate.

Later work in the laboratory of the inventors identified the full sequence, set forth above. Fragments of this total sequence can readily be prepared using restriction endonucleases. Computer analysis of both the forward and reverse sequence has identified a number of cleavage sites.

### III. ET-NANB Fragments

According to another aspect, the invention includes ET-NANB-specific fragments or probes which hybridize with ET-NANB genomic sequences or cDNA fragments derived therefrom. The fragments may include full-length cDNA fragments such as described in

Section II, or may be derived from shorter sequence regions within cloned cDNA fragments. Shorter fragments can be prepared by enzymatic digestion of full-length fragments under conditions which yield desired-sized fragments, as will be described in Section IV. Alternatively, the fragments can be produced by oligonucleotide synthetic methods, using sequences derived from the cDNA fragments. Methods or commercial services for producing selected-sequence oligonucleotide fragments are available. Fragments are usually at least 12 nucleotides in length, preferably at least 14, 20, 30 or 50 nucleotides, when used as probes. Probes can be full length or less than 500, preferably less than 300 or 200, nucleotides in length.

To confirm that a given ET-NANB fragment is in fact derived from the ET-NANB viral agent, the fragment can be shown to hybridize selectively with cDNA from infected sources. By way of illustration, to confirm that the 1.33 kb fragment in the pTZKF1(ET1.1) plasmid is ET-NANB in origin, the fragment was excised from the pTZKF1(ET1.1) plasmid, purified, and radiolabeled by random labeling. The radiolabeled fragment was hybridized with fractionated cDNAs from infected and non-infected sources to confirm that the probe reacts only with infected-source cDNAs. This method is illustrated in Example 4, where the above radiolabeled 1.33 kb fragment from pTZKF1(ET1.1) plasmid was examined for binding to cDNAs prepared from infected and non-infected sources. The infected sources are (1) bile from a cynomolgus macaque infected with a strain of virus derived from stool samples from human patients from Burma with known ET-NANB infections and (2) a viral agent derived from the stool sample of a human ET-NANB patient from Mexico. The cDNAs in each fragment mixture were first amplified by a linker primer amplification method described in Example 4. Fragment separation was on

agarose gel, followed by Southern blotting and then hybridization to bind the radiolabeled 1.33 kb fragment to the fractionated cDNAs. The lane containing cDNAs from the infected sources showed a smeared band of bound probe, as expected (cDNAs amplified by the linker/primer amplification method would be expected to have a broad range of sizes). No probe binding to the amplified cDNAs from the non-infected sources was observed. The results indicate that the 1.33 kb probe is specific for cDNA fragments associated with ET-NANB infection. This same type of study, using ET 1.1 as the probe, has demonstrated hybridization to ET-NANB samples collected from Tashkent, Somalia, Borneo and Pakistan. Secondly, the fact that the probe is specific for ET-NANB related sequences derived from different continents (Asia, Africa and North America) indicates the cloned ET-NANB Burma sequence (ET1.1) is derived from a common ET-NANB virus or virus class responsible for ET-NANB hepatitis infection worldwide.

In a related confirmatory study, probe binding to fractionated genomic fragments prepared from human or cynomolgus macaque genomic DNA (both infected and uninfected) was examined. No probe binding was observed to either genomic fraction, demonstrating that the ET-NANB fragment is not an endogenous human or cynomolgus genomic fragment and additionally demonstrating that HEV is an RNA virus.

Another confirmation of ET-NANB specific sequences in the fragments is the ability to express ET-NANB proteins from coding regions in the fragments and to demonstrated specific sero-reactivity of these proteins with sera collected during documented outbreaks of ET-NANB. Section IV below discusses methods of protein expression using the fragments.

One important use of the ET-NANB-specific fragments is for identifying ET-NANB-derived cDNAs which contain additional sequence information. The

newly identified cDNAs, in turn, yield new fragment probes, allowing further iterations until the entire viral genome is identified and sequenced. Procedures for identifying additional ET-NANB library clones and  
5 generating new probes therefrom generally follow the cloning and selection procedures described in Section II.

The fragments (and oligonucleotides prepared based on the sequences given above) are also useful as  
10 primers for a polymerase chain reaction method of detecting ET-NANB viral genomic material in a patient sample. This diagnostic method will be described in Section V below.

Two specific genetic sequences derived from  
15 the Mexican strain, identified herein as 406.3-2 and 406.4-2, have been identified that encode immunogenic epitopes. This was done by isolating clones which encode epitopes that immunologically react specifically with sera from individuals and  
20 experimental animals infected with HEV. Comparison of the isolated sequences with those in the Genebank collection of genetic sequences indicate that these viral sequences are novel. Since these sequences are unique, they can be used to identify the presence of  
25 HEV and to distinguish this strain of hepatitis from HAV, HBV, and HCV strains. The sequences are also useful for the design of oligonucleotide probes to diagnose the presence of virus in samples. They can be used for the synthesis of polypeptides that  
30 themselves are used in immunoassays. The specific 406.3-2 and 406.4-2 sequences can be incorporated into other genetic material, such as vectors, for ease of expression or replication. They can also be used (as demonstrated above) for identifying similar antigenic  
35 regions encoded by related viral strains, such as the Burmese strain.

#### IV. ET-NANB Proteins

As indicated above, ET-NANB proteins can be prepared by expressing open reading-frame coding regions in ET-NANB fragments. In one preferred approach, the ET-NANB fragments used for protein expression are derived from cloned cDNAs which have been treated to produce desired-size fragments, and preferably random fragments with sizes predominantly between about 100 to about 300 base pairs. Example 5 describes the preparation of such fragments by DNAs digestion. Because it is desired to obtain peptide antigens of between about 30 to about 100 amino acids, the digest fragments are preferably size fractionated, for example by gel electrophoresis, to select those in the approximately 100-300 basepair size range. Alternatively, cDNA libraries constructed directly from HEV-containing sources (e.g., bile or stool) can be screened directly if cloned into an appropriate expression vector (see below).

For example, the ET-NANB proteins expressed by the 406.3-2 and 406.4-2 sequences (and peptide fragments thereof) are particularly preferred since these proteins have been demonstrated to be immunoreactive with a variety of different human sera, thereby indicating the presence of one or more epitopes specific for HEV on their surfaces. These clones were identified by direct screening of a gt11 library.

#### A. Expression Vector

The ET-NANB fragments are inserted into a suitable expression vector. One exemplary expression vector is lambda gt11, which contains a unique EcoRI insertion site 53 base pairs upstream of the translation termination codon of the beta-galactosidase gene. Thus, the inserted sequence will be expressed as a beta-galactosidase fusion protein which contains the N-terminal portion of the beta-galactosidase gene, the heterologous peptide, and

optionally the C-terminal region of the beta-galactosidase peptide (the C-terminal portion being expressed when the heterologous peptide coding sequence does not contain a translation termination codon). This vector also produces a temperature-sensitive repressor (c1857) which causes viral lysogeny at permissive temperatures, e.g., 32°C, and leads to viral lysis at elevated temperatures, e.g., 37°C. Advantages of this vector include: (1) highly efficient recombinant generation, (2) ability to select lysogenized host cells on the basis of host-cell growth at permissive, but not non-permissive, temperatures, and (3) high levels of recombinant fusion protein production. Further, since phage containing a heterologous insert produces an inactive beta-galactosidase enzyme, phage with inserts can be readily identified by a beta-galactosidase colored-substrate reaction.

For insertion into the expression vector, the viral digest fragments may be modified, if needed, to contain selected restriction-site linkers, such as EcoRI linkers, according to conventional procedures. Example 1 illustrates methods for cloning the digest fragments into lambda gt11, which includes the steps of blunt-ending the fragments, ligating with EcoRI linkers, and introducing the fragments into EcoRI-cut lambda gt11. The resulting viral genomic library may be checked to confirm that a relatively large (representative) library has been produced. This can be done, in the case of the lambda gt11 vector, by infecting a suitable bacterial host, plating the bacteria, and examining the plaques for loss of beta-galactosidase activity. Using the procedures described in Example 1, about 50% of the plaques showed loss of enzyme activity.

#### B. Peptide Antigen Expression

The viral genomic library formed above is screened for production of peptide antigen (expressed as a fusion protein) which is immunoreactive with antiserum from ET-NANB seropositive individuals . In  
5 a preferred screening method, host cells infected with phage library vectors are plated, as above, and the plate is blotted with a nitrocellulose filter to transfer recombinant protein antigens produced by the cells onto the filter. The filter is then reacted with  
10 the ET-NANB antiserum, washed to remove unbound antibody, and reacted with reporter-labeled, anti-human antibody, which becomes bound to the filter, in sandwich fashion, through the anti-ET-NANB antibody.

Typically phage plaques which are identified  
15 by virtue of their production of recombinant antigen of interest are re-examined at a relatively low density for production of antibody-reactive fusion protein. Several recombinant phage clones which produced immunoreactive recombinant antigen were  
20 identified in the procedure.

The selected expression vectors may be used for scale-up production, for purposes of recombinant protein purification. Scale-up production is carried out using one of a variety of reported methods for (a)  
25 lysogenizing a suitable host, such as E. coli, with a selected lambda gt11 recombinant (b) culturing the transduced cells under conditions that yield high levels of the heterologous peptide, and (c) purifying the recombinant antigen from the lysed cells.

30 In one preferred method involving the above lambda gt11 cloning vector, a high-producer E. coli host, BNN103, is infected with the selected library phage and replica plated on two plates. One of the plates is grown at 32°C, at which viral lysogeny can  
35 occur, and the other at 42°C, at which the infecting phage is in a lytic stage and therefore prevents cell growth. Cells which grow at the lower but not the

higher temperature are therefore assumed to be successfully lysogenized.

The lysogenized host cells are then grown under liquid culture conditions which favor high  
5 production of the fused protein containing the viral insert, and lysed by rapid freezing to release the desired fusion protein.

#### C. Peptide Purification

10 The recombinant peptide can be purified by standard protein purification procedures which may include differential precipitation, molecular sieve chromatography, ion-exchange chromatography,  
isoelectric focusing, gel electrophoresis and  
15 affinity chromatography. In the case of a fused protein, such as the beta-galactosidase fused protein prepared as above, the protein isolation techniques which are used can be adapted from those used in isolation of the native protein. Thus, for isolation  
20 of a soluble betagalactosidase fusion protein, the protein can be isolated readily by simple affinity chromatography, by passing the cell lysis material over a solid support having surface-bound anti-beta-galactosidase antibody.

25

#### D. Viral Proteins

The ET-NANB protein of the invention may also be derived directly from the ET-NANB viral agent. VLPs or protein isolated from stool or liver samples  
30 from an infected individual, as above, are one suitable source of viral protein material. The VLPs isolated from the stool sample may be further purified by affinity chromatography prior to protein isolation (see below). The viral agent may also be raised in  
35 cell culture, which provides a convenient and potentially concentrated source of viral protein. Co-owned U.S. Patent Application Serial No. 846,757, filed April 1, 1986, describes an immortalized trioma

liver cell which supports NANB infection in cell culture. The trioma cell line is prepared by fusing human liver cells with a mouse/human fusion partner selected for human chromosome stability. Cells  
5 containing the desired NANB viral agent can be identified by immunofluorescence methods, employing anti-ET-NANB human antibodies.

The viral agent is disrupted, prior to protein isolation, by conventional methods, which can  
10 include sonication, high- or low-salt conditions, or use of detergents.

Purification of ET-NANB viral protein can be carried out by affinity chromatography, using a purified anti-ET-NANB antibody attached according to  
15 standard methods to a suitable solid support. The antibody itself may be purified by affinity chromatography, where an immunoreactive recombinant ETNANB protein, such as described above, is attached to a solid support, for isolation of anti-ET-NANB  
20 antibodies from an immune serum source. The bound antibody is released from the support by standard methods.

Alternatively, the anti-ET-NANB antibody may be an antiserum or a monoclonal antibody (Mab)  
25 prepared by immunizing a mouse or other animal with recombinant ETNANB protein. For Mab production, lymphocytes are isolated from the animal and immortalized with a suitable fusion partner, and successful fusion products which react with the  
30 recombinant protein immunogen are selected. These in turn may be used in affinity purification procedures, described above, to obtain native ET-NANB antigen.

#### V. Utility

35 Although ET-NANB is primarily of interest because of its effects on humans, recent data has shown that this virus is also capable of infecting other animals, especially mammals. Accordingly, any

discussion herein of utility applies to both human and veterinary uses, especially commercial veterinary uses, such as the diagnosis and treatment of pigs, cattle, sheep, horses, and other domesticated animals.

5           A.    Diagnostic Methods

          The particles and antigens of the invention, as well as the genetic material, can be used in diagnostic assays. Methods for detecting the presence of ET-NANB hepatitis comprise analyzing a biological  
10   sample such as a blood sample, stool sample or liver biopsy specimen for the presence of an analyte associated with ET-NANB hepatitis virus.

          The analyte can be a nucleotide sequence which hybridizes with a probe comprising a sequence of  
15   at least about 16 consecutive nucleotides, usually 30 to 200 nucleotides, up to substantially the full sequence of the sequences shown above (cDNA sequences). The analyte can be RNA or cDNA. The analyte is typically a virus particle suspected of  
20   being ET-NANB or a particle for which this classification is being ruled out. The virus particle can be further characterized as having an RNA viral genome comprising a sequence at least about 70% homologous to a sequence of at least 12 consecutive  
25   nucleotides of the "forward" and "reverse" sequences given above, usually at least about 80% homologous to at least about 60 consecutive nucleotides within the sequences, and may comprise a sequence substantially homologous to the full-length sequences. In order to  
30   detect an analyte, where the analyte hybridizes to a probe, the probe may contain a detectable label. Particularly preferred for use as a probe are sequences of consecutive nucleotides derived from the 406.3-2 and 406.4-2 clones described herein, since  
35   these clones appear to be particularly diagnostic for HEV.

          The analyte can also comprise an antibody which recognizes an antigen, such as a cell surface

antigen, on a ET-NANB virus particle. The analyte can also be a ET-NANB viral antigen. Where the analyte is an antibody or an antigen, either a labelled antigen or antibody, respectively, can be used to bind to the analyte to form an immunological complex, which can then be detected by means of the label.

Typically, methods for detecting analytes such as surface antigens and/or whole particles are based on immunoassays. Immunoassays can be conducted either to determine the presence of antibodies in the host that have arisen from infection by ET-NANB hepatitis virus or by assays that directly determine the presence of virus particles or antigens. Such techniques are well known and need not be described here in detail. Examples include both heterogeneous and homogeneous immunoassay techniques. Both techniques are based on the formation of an immunological complex between the virus particle or its antigen and a corresponding specific antibody. Heterogeneous assays for viral antigens typically use a specific monoclonal or polyclonal antibody bound to a solid surface. Sandwich assays are becoming increasingly popular. Homogeneous assays, which are carried out in solution without the presence of a solid phase, can also be used, for example by determining the difference in enzyme activity brought on by binding of free antibody to an enzyme-antigen conjugate. A number of suitable assays are disclosed in U.S. Patent Nos. 3,817,837, 4,006,360, 3,996,345.

When assaying for the presence of antibodies induced by ET-NANB viruses, the viruses and antigens of the invention can be used as specific binding agents to detect either IgG or IgM antibodies. Since IgM antibodies are typically the first antibodies that appear during the course of an infection, when IgG synthesis may not yet have been initiated, specifically distinguishing between IgM and IgG antibodies present in the blood stream of a host will

enable a physician or other investigator to determine whether the infection is recent or convalescent.

Proteins expressed by the 406.3-2 and 406.4-2 clones described herein and peptide fragments thereof are particularly preferred for use as specific binding agents to detect antibodies since they have been demonstrated to be reactive with a number of different human HEV sera. Further, they are reactive with both acute and convalescent sera.

10           In one diagnostic configuration, test serum is reacted with a solid phase reagent having surface-bound ET-NANB protein antigen. After binding anti-ET-NANB antibody to the reagent and removing unbound serum components by washing, the reagent is reacted with reporter-labeled anti-human antibody to bind reporter to the reagent in proportion to the amount of bound anti-ET-NANB antibody on the solid support. The reagent is again washed to remove unbound labeled antibody, and the amount of reporter associated with the reagent is determined. Typically, the reporter is an enzyme which is detected by incubating the solid phase in the presence of a suitable fluorometric or colorimetric substrate.

25           The solid surface reagent in the above assay prepared by known techniques for attaching protein material to solid support material, such as polymeric beads, dip sticks, or filter material. These attachment methods generally include non-specific adsorption of the protein to the support or covalent attachment of the protein, typically through a free amine group, to a chemically reactive group on the solid support, such as an activate carboxyl, hydroxyl, or aldehyde group.

35           In a second diagnostic configuration, known as a homogeneous assay, antibody binding to a solid support produces some change in the reaction medium which can be directly detected in the medium. Known general types of homogeneous assays proposed

heretofore include (a) spin-labeled reporters, where antibody binding to the antigen is detected by a change in reported mobility (broadening of the spin splitting peaks), (b) fluorescent reporters, where  
5 binding is detected by a change in fluorescence efficiency, (c) enzyme reporters, where antibody binding effects enzyme/substrate interactions, and (d) liposome-bound reporters, where binding leads to liposome lysis and release of encapsulated reporter.  
10 The adaptation of these methods to the protein antigen of the present invention follows conventional methods for preparing homogeneous assay reagents.

In each of the assays described above, the assay method involves reacting the serum from a test  
15 individual with the protein antigen and examining the antigen for the presence of bound antibody. The examining may involve attaching a labeled anti-human antibody to the antibody being examined, either IgM (acute phase) or IgG (convalescent phase), and  
20 measuring the amount of reporter bound to the solid support, as in the first method, or may involve observing the effect of antibody binding on a homogeneous assay reagent, as in the second method.

Also forming part of the invention is an  
25 assay system or kit for carrying out the assay method just described. The kit generally includes a support with surface-bound recombinant protein antigen which is (a) immunoreactive with antibodies present in individuals infected with enterically transmitted  
30 nonA/nonB viral agent and (b) derived from a viral hepatitis agent whose genome contains a region which is homologous to the 1.33 kb DNA EcoRI insert present in plasmid pTZKF1(ET1.1) carried in E. Coli strain BB4, and having ATCC deposit no. 67717. A reporter-labeled anti-human antibody in the kit is used for  
35 detecting surface-bound anti-ET-NANB antibody.

## B. Viral Genome Diagnostic Applications

The genetic material of the invention can itself be used in numerous assays as probes for genetic material present in naturally occurring infections. One method for amplification of target nucleic acids, for later analysis by hybridization assays, is known as the polymerase chain reaction or PCR technique. The PCR technique can be applied to detecting virus particles of the invention in suspected pathological samples using oligonucleotide primers spaced apart from each other and based on the genetic sequence set forth above. The primers are complementary to opposite strands of a double stranded DNA molecule and are typically separated by from about 50 to 450 nt or more (usually not more than 2000 nt). This method entails preparing the specific oligonucleotide primers and then repeated cycles of target DNA denaturation, primer binding, and extension with a DNA polymerase to obtain DNA fragments of the expected length based on the primer spacing. Extension products generated from one primer serve as additional target sequences for the other primer. The degree of amplification of a target sequence is controlled by the number of cycles that are performed and is theoretically calculated by the simple formula  $2^n$  where  $n$  is the number of cycles. Given that the average efficiency per cycle ranges from about 65% to 85%, 25 cycles produce from 0.3 to 4.8 million copies of the target sequence. The PCR method is described in a number of publications, including Saiki *et al.*, Science (1985) 230:1350-1354; Saiki *et al.*, Nature (1986) 324:163-166; and Scharf *et al.*, Science (1986) 233:1076-1078. Also see U.S. Patent Nos. 4,683,194; 4,683,195; and 4,683,202.

The invention includes a specific diagnostic method for determination of ET-NANB viral agent, based on selective amplification of ET-NANB fragments. This method employs a pair of single-strand primers derived

from non-homologous regions of opposite strands of a DNA duplex fragment, which in turn is derived from an enterically transmitted viral hepatitis agent whose genome contains a region which is homologous to the  
5 1.33 kb DNA EcoRI insert present in plasmid pTZKF1(ET1.1) carried in E. coli strain BB4, and having ATCC deposit no. 67717. These "primer fragments," which form one aspect of the invention, are prepared from ET-NANB fragments such as described  
10 in Section III above. The method follows the process for amplifying selected nucleic acid sequences as disclosed in U.S. Patent No. 4,683,202, as discussed above.

15 C. Peptide Vaccine

Any of the antigens of the invention can be used in preparation of a vaccine. A preferred starting material for preparation of a vaccine is the particle antigen isolated from bile. The antigens are  
20 preferably initially recovered as intact particles as described above. However, it is also possible to prepare a suitable vaccine from particles isolated from other sources or non-particle recombinant antigens. When non-particle antigens are used (typically soluble  
25 antigens), proteins derived from the viral envelope or viral capsid are preferred for use in preparing vaccines. These proteins can be purified by affinity chromatography, also described above.

If the purified protein is not immunogenic  
30 per se, it can be bound to a carrier to make the protein immunogenic. Carriers include bovine serum albumin, keyhole limpet hemocyanin and the like. It is desirable, but not necessary, to purify antigens to be substantially free of human protein. However, it is  
35 more important that the antigens be free of proteins, viruses, and other substances not of human origin that may have been introduced by way of, or contamination of, the nutrient medium, cell lines,

tissues, or pathological fluids from which the virus is cultured or obtained.

Vaccination can be conducted in conventional fashion. For example, the antigen, whether a viral particle or a protein, can be used in a suitable diluent such as water, saline, buffered salines, complete or incomplete adjuvants, and the like. The immunogen is administered using standard techniques for antibody induction, such as by subcutaneous administration of physiologically compatible, sterile solutions containing inactivated or attenuated virus particles or antigens. An immune response producing amount of virus particles is typically administered per vaccinizing injection, typically in a volume of one milliliter or less.

A specific example of a vaccine composition includes, in a pharmacologically acceptable adjuvant, a recombinant protein or protein mixture derived from an enterically transmitted nonA/nonB viral hepatitis agent whose genome contains a region which is homologous to the 1.33 kb DNA EcoRI insert present in plasmid pTZKF1(ET1.1) carried in E. coli strain BB4, and having ATCC deposit no. 67717. The vaccine is administered at periodic intervals until a significant titer of anti-ET-NANB antibody is detected in the serum. The vaccine is intended to protect against ET-NANB infection.

Particularly preferred are vaccines prepared using proteins expressed by the 406.3-2 and 406.4-2 clones described herein and equivalents thereof, including fragments of the expressed proteins. Since these clones have already been demonstrated to be reactive with a variety of human HEV-positive sera, their utility in protecting against a variety of HEV strains is indicated.

D. Prophylactic and Therapeutic  
Antibodies and Antisera

In addition to use as a vaccine, the compositions can be used to prepare antibodies to ET-NANB virus particles. The antibodies can be used directly as antiviral agents. To prepare antibodies, a host animal is immunized using the virus particles or, as appropriate, non-particle antigens native to the virus particle are bound to a carrier as described above for vaccines. The host serum or plasma is collected following an appropriate time interval to provide a composition comprising antibodies reactive with the virus particle. The gamma globulin fraction or the IgG antibodies can be obtained, for example, by use of saturated ammonium sulfate or DEAE Sephadex, or other techniques known to those skilled in the art. The antibodies are substantially free of many of the adverse side effects which may be associated with other anti-viral agents such as drugs.

The antibody compositions can be made even more compatible with the host system by minimizing potential adverse immune system responses. This is accomplished by removing all or a portion of the FC portion of a foreign species antibody or using an antibody of the same species as the host animal, for example, the use of antibodies from human/human hybridomas.

The antibodies can also be used as a means of enhancing the immune response since antibody-virus complexes are recognized by macrophages. The antibodies can be administered in amounts similar to those used for other therapeutic administrations of antibody. For example, pooled gamma globulin is administered at 0.02-0.1 ml/lb body weight during the early incubation of other viral diseases such as rabies, measles and hepatitis B to interfere with viral entry into cells. Thus, antibodies reactive with the ET-NANB virus particle can be passively administered alone or in conjunction with another anti-viral agent to a host infected with an ET-NANB virus to enhance the immune

response and/or the effectiveness of an antiviral drug.

Alternatively, anti-ET-NANB-virus antibodies can be induced by administering anti-idiotypic anti-  
5 bodies as immunogens. Conveniently, a purified anti--  
ET-NANB-virus antibody preparation prepared as de-  
scribed above is used to induce anti-idiotypic antibody  
in a host animal. The composition is administered to  
10 the host animal in a suitable diluent. Following  
administration, usually repeated administration, the  
host produces anti-idiotypic antibody. To eliminate an  
immunogenic response to the Fc region, antibodies pro-  
duced by the same species as the host animal can be  
15 used or the Fc region of the administered antibodies  
can be removed. Following induction of anti-idiotypic  
antibody in the host animal, serum or plasma is  
removed to provide an antibody composition. The  
composition can be purified as described above for  
20 anti-ET-NANB virus antibodies, or by affinity  
chromatography using anti-ET-NANB-virus antibodies  
bound to the affinity matrix. The anti-idiotypic  
antibodies produced are similar in conformation to the  
authentic ET-NANB antigen and may be used to prepare  
an ET-NANB vaccine rather than using a ET-NANB  
25 particle antigen.

When used as a means of inducing anti-ET-  
NANB virus antibodies in a patient, the manner of  
injecting the antibody is the same as for vaccination  
purposes, namely intramuscularly, intraperitoneally,  
30 subcutaneously or the like in an effective  
concentration in a physiologically suitable diluent  
with or without adjuvant. One or more booster  
injections may be desirable. The anti-idiotypic method  
of induction of anti-ET-NANB virus antibodies can  
35 alleviate problems which may be caused by passive  
administration of anti-ET-NANB-virus antibodies, such  
as an adverse immune response, and those associated

with administration of purified blood components, such as infection with as yet undiscovered viruses.

5 The ET-NANB derived proteins of the invention are also intended for use in producing antiserum designed for pre- or post-exposure prophylaxis. Here an ET-NANB protein, or mixture of proteins is formulated with a suitable adjuvant and administered by injection to human volunteers, according to known methods for producing human antisera. Antibody response to the  
10 injected proteins is monitored, during a several- week period following immunization, by periodic serum sampling to detect the presence an anti-ET-NANB serum antibodies, as described in Section IIA above.

The antiserum from immunized individuals may  
15 be administered as a pre-exposure prophylactic measure for individuals who are at risk of contracting infection. The antiserum is also useful in treating an individual post-exposure, analogous to the use of high titer antiserum against hepatitis B virus for post-  
20 exposure prophylaxis.

#### E. Monoclonal Antibodies

For both in vivo use of antibodies to ET-NANB virus particles and proteins and anti-idiotypic  
25 antibodies and diagnostic use, it may be preferable to use monoclonal antibodies. Monoclonal anti-virus particle antibodies or anti-idiotypic antibodies can be produced as follows. The spleen or lymphocytes from an immunized animal are removed and immortalized or used  
30 to prepare hybridomas by methods known to those skilled in the art. To produce a human-human hybridoma, a human lymphocyte donor is selected. A donor known to be infected with a ET-NANB virus (where infection has been shown for example by the presence  
35 of anti-virus antibodies in the blood or by virus culture) may serve as a suitable lymphocyte donor. Lymphocytes can be isolated from a peripheral blood sample or spleen cells may be used if the donor is

subject to splenectomy. Epstein-Barr virus (EBV) can be used to immortalize human lymphocytes or a human fusion partner can be used to produce human-human hybridomas. Primary in vitro immunization with  
5 peptides can also be used in the generation of human monoclonal antibodies.

Antibodies secreted by the immortalized cells are screened to determine the clones that secrete antibodies of the desired specificity. For  
10 monoclonal anti-virus particle antibodies, the antibodies must bind to ET-NANB virus particles. For monoclonal anti-idiotypic antibodies, the antibodies must bind to anti-virus particle antibodies. Cells producing antibodies of the desired specificity are  
15 selected.

The following examples illustrate various aspects of the invention, but are in no way intended to limit the scope thereof.

#### 20 Material

The materials used in the following Examples were as follows:

Enzymes: DNase I and alkaline phosphatase were obtained from Boehringer Mannheim Biochemicals  
25 (BMB, Indianapolis, IN); EcoRI, EcoRI methylase, DNA ligase, and DNA Polymerase I, from New England Biolabs (NEB, Beverly MA); and RNase A was obtained from Sigma (St. Louis, MO) .

Other reagents: EcoRI linkers were obtained  
30 from NEB; and nitro blue tetrazolium (NBT), S-bromo-4-chloro-3-indolyl phosphate (BCIP) S-bromo-4-chloro-3-indolyl-B-D-galactopyranoside (Xgal) and isopropyl B-D-thiogalactopyranoside (IPTG) were obtained from Sigma.

35 cDNA synthesis kit and random priming labeling kits are available from Boehringer-Mannheim Biochemical (BMB, Indianapolis, IN).

Example 1  
Preparing cDNA Library

A. Source of ET-NANB virus

Two cynomolgus monkeys (cynos) were  
5 intravenously injected with a 10% suspension of a  
stool pool obtained from a second-passage cyno (cyno  
#37) infected with a strain of ET-NANB virus isolated  
from Burma cases whose stools were positive for ET-  
NANB, as evidenced by binding of 27-34 nm virus-like  
10 particles (VLPs) in the stool to immune serum from a  
known ETNANB patient. The animals developed elevated  
levels of alanine aminotransferase (ALT) between 24-36  
days after inoculation, and one excreted 27-34 nm  
VLPs in its bile in the pre-acute phase of infection.

15 The bile duct of each infected animal was  
cannulated and about 1-3 cc of bile was collected  
daily. RNA was extracted from one bile specimen (cyno  
#121) by hot phenol extraction, using a standard RNA  
isolation procedure. Double-strand cDNA was formed  
20 from the isolated RNA by a random primer for first-  
strand generation, using a cDNA synthesis kit obtained  
from Boehringer-Mannheim (Indianapolis, IN).

B. Cloning the Duplex Fragments

25 The duplex cDNA fragments were blunt-ended  
with T4 DNA polymerase under standard conditions  
(Maniatis, p. 118), then extracted with  
phenol/chloroform and precipitated with ethanol. The  
blunt-ended material was ligated with EcoRI linkers  
30 under standard conditions (Maniatis, pp. 396-397) and  
digested with EcoRI to remove redundant linker ends.  
Non-ligated linkers were removed by sequential  
isopropanol precipitation.

Lambda gt10 phage vector (Huynh) was  
35 obtained from Promega Biotec (Madison, WI). This  
cloning vector has a unique EcoRI cloning site in the  
phage CI repressor gene. The cDNA fragments from above  
were introduced into the EcoRI site by mixing 0.5 -

1.0  $\mu$ g EcoRI-cleaved gt10, 0.5-3  $\mu$ l of the above  
duplex fragments, 0.5  $\mu$ l 10X ligation buffer, 0.5  $\mu$ l  
ligase (200 units), and distilled water to 5  $\mu$ l. The  
mixture was incubated overnight at 14°C, followed by  
5 in vitro packaging, according to standard methods  
(Maniatis, pp. 256-268).

The packaged phage were used to infect an E.  
coli hfl strain, such as strain HG415. Alternatively,  
E. coli, strain C600 hfl available from Promega  
10 Biotec, Madison, WI, could be used. The percentage of  
recombinant plaques obtained with insertion of the  
EcoRI-ended fragments was less than 5% by analysis of  
20 random plaques.

The resultant cDNA library was plated and  
15 phage were eluted from the selection plates by  
addition of elution buffer. After DNA extraction from  
the phage, the DNA was digested with EcoRI to release  
the heterogeneous insert population, and the DNA  
fragments were fractionated on agarose to remove phage  
20 fragments. The 500-4,000 basepair inserts were  
isolated and recloned into lambda gt10 as above, and  
the packaged phage was used to infect E. coli strain  
HG415. The percentage of successful recombinants was  
greater than 95%. The phage library was plated on E.  
25 coli strain HG415, at about 5,000 plaques/plate, on a  
total of 8 plates.

## Example 2

### Selecting ET-NANB Cloned Fragments

#### 30 A. cDNA Probes

Duplex cDNA fragments from noninfected and  
ETNANB-infected cynomolgus monkeys were prepared as in  
Example 1. The cDNA fragments were radiolabeled by  
random priming, using a random-priming labeling kit  
35 obtained from Boehringer-Mannheim (Indianapolis, IN).

#### B. Clone Selection

The plated cDNA library from Example 1 was transferred to each of two nitrocellulose filters, and the phage DNA was fixed on the filters by baking, according to standard methods (Maniatis, pp. 320323).  
5 The duplicate filters were hybridized with either infected-source or control cDNA probes from above. Autoradiographs of the filters were examined to identify library clones which hybridized with radiolabeled cDNA probes from infected source only,  
10 i.e., did not hybridize with cDNA probes from the non-infected source. Sixteen such clones, out of a total of about 40,000 clones examined, were identified by this subtraction selection method.

Each of the sixteen clones was picked and  
15 replated at low concentration on an agar plate. The clones on each plate were transferred to two nitrocellulose ag duplicate lifts, and examined for hybridization to radiolabeled cDNA probes from infected and noninfected sources, as above. Clones were selected  
20 which showed selective binding for infected-source probes (i.e., binding with infected-source probes and substantially no binding with non-infected-source probes). One of the clones which bound selectively to probe from infected source was isolated for further  
25 study. The selected vector was identified as lambda gt10-1.1, indicated in Figure 1.

### Example 3

#### ET-NANB Sequence

30 Clone lambda gt10-1.1 from Example 2 was digested with EcoRI to release the heterologous insert, which was separated from the vector fragments by gel electrophoresis. The electrophoretic mobility of the fragment was consistent with a 1.33 kb fragment. This  
35 fragment, which contained EcoRI ends, was inserted into the EcoRI site of a pTZKF1 vector, whose construction and properties are described in co-owned U.S. patent application for "Cloning Vector System and

Method for Rare Clone Identification", Serial No. 125, 650, filed November 25, 1987. Briefly, and as illustrated in Figure 1, this plasmid contains a unique EcoRI site adjacent a T7 polymerase promoter site, and plasmid and phage origins of replication. The sequence immediately adjacent each side of the EcoRI site is known. E. coli BB4 bacteria, obtained from Stratagene (La Jolla, CA, were transformed with the plasmid.

Radiolabeled ET-NANB probe was prepared by excising the 1.33 kb insert from the lambda gt10-1.1 phage in Example 2, separating the fragment by gel electrophoresis, and randomly labeling as above. Bacteria transfected with the above pTZKF1 and containing the desired ET-NANB insert were selected by replica lift and hybridization with the radiolabeled ET-NANB probe, according to methods outlined in Example 2.

One bacterial colony containing a successful recombinant was used for sequencing a portion of the 1.33 kb insert. This isolate, designated pTZKF1(ET1.1), has been deposited with the American Type Culture Collection, and is identified by ATCC deposit no. 67717. Using a standard dideoxy sequencing procedure, and primers for the sequences flanking the EcoRI site, about 200-250 basepairs of sequence from the 5'-end region and 3'-end region of the insert were obtained. The sequences are given above in Section II. Later sequencing by the same techniques gave the full sequence in both directions, also given above.

#### Example 4

##### Detecting ET-NANB Sequences

cDNA fragment mixtures from the bile of noninfected and ET-NANB-infected cynomolgus monkeys were prepared as above. The cDNA fragments obtained from human stool samples were prepared as follows.

This ml of a 10% stool suspension obtained from an individual from Mexico diagnosed as infected with ET-NANB as a result of an ET-NANB outbreak, and a similar volume of stool from a healthy, non-infected

5 individual, were layered over a 10% sucrose density gradient cushion, and centrifuged at 25,000 x g for 6 hr in an SW27 rotor, at 18°C. The pelleted material from the infected-source stool contained 27-34 nm VLP particles characteristic of ET-NANB infection in the  
10 infected-stool sample. RNA was isolated from the sucrose-gradient pellets in both the infected and non-infected samples, and the isolated RNA was used to produce cDNA fragments as described in Example 1.

The cDNA fragment mixtures from infected and  
15 non-infected bile source, and from infected and non-infected human-stool source were each amplified by a novel linker/primer replication method described in co-owned patent application serial number 07/208,512 for "DNA Amplification and Subtraction Technique,"  
20 filed June 17, 1988. Briefly, the fragments in each sample were blunt-ended with DNA Pol I then extracted with phenol/chloroform and precipitated with ethanol. The blunt-ended material was ligated with linkers having the following sequence (top or 5' sequence has  
25 SEQ ID NO.21; bottom or 3' sequence has SEQ ID NO:22):

5'-GGAATTCGCGGCCGCTCG-3'

3'-TTCCTTAAGCGCCGCGGAGC-5'

The duplex fragments were digested with  
30 NruI to remove linker dimers, mixed with a primer having the sequence 5'-GGAATTCGCGGCCGCTCG-3', and then heat denatured and cooled to room temperature to form single-strand DNA/primer complexes. The complexes were replicated to form duplex fragments by addition of  
35 *Thermus aquaticus* (Taq) polymerase and all four deoxynucleotides. The replication procedures, involving successive strand denaturation, formation of

strand/primer complexes, and replication, was repeated 25 times.

The amplified cDNA sequences were fractionated by agarose gel electrophoresis, using a 2% agarose matrix. After transfer of the DNA fragments from the agarose gels to nitrocellulose paper, the filters were hybridized to a random-labeled 32p probe prepared by (i) treating the pTZKF1(ET1.1) plasmid from above with EcoRI, (ii) isolating the released 1.33 kb ET-NANB fragment, and (iii) randomly labeling the isolated fragment. The probe hybridization was performed by conventional Southern blotting methods (Maniatis, pp. 382-389). Figure 2 shows the hybridization pattern obtained with cDNAs from infected (I) and non-infected (N) bile sources (2A) and from infected (I) and noninfected (N) human stool sources (2B). As seen, the ET-NANB probe hybridized with fragments obtained from both of the infected sources, but was non-homologous to sequences obtained from either of the non-infected sources, thus confirming the specificity of derived sequence.

Southern blots of the radiolabeled 1.33 kb fragment with genomic DNA fragments from both human and cynomolgus-monkey DNA were also prepared. No probe hybridization to either of the genomic fragment mixtures was observed, confirming that the ET-NANB sequence is exogenous to either human or cynomolgus genome.

30

#### Example 5

##### Expressing ET-NANB Proteins

##### A. Preparing ET-NANB Coding Sequences

The pTZKF1(ET1.1) plasmid from Example 2 was digested with EcoRI to release the 1.33 kb ET-NANB insert which was purified from the linearized plasmid by gel electrophoresis. The purified fragment was suspended in a standard digest buffer (0.5M Tris HCl, pH 7.5; 1 mg/ml BSA; 10mM MnCl<sub>2</sub>) to a concentration of

about 1 mg/ml and digested with DNase I at room temperature for about 5 minutes. These reaction conditions were determined from a prior calibration study, in which the incubation time required to produce predominantly 100-300 basepair fragments was determined. The material was extracted with phenol/chloroform before ethanol precipitation.

The fragments in the digest mixture were blunt-ended and ligated with EcoRI linkers as in Example 1. The resultant fragments were analyzed by electrophoresis (5-10V/cm) on 1.2% agarose gel, using PhiX174/HaeIII and lambda/HindIII size markers. The 100-300 bp fraction was eluted onto NA45 strips (Schleicher and Schuell), which were then placed into 1.5 ml microtubes with eluting solution (1 M NaCl, 50 mM arginine, pH 9.0), and incubated at 67°C for 30-60 minutes. The eluted DNA was phenol/chloroform extracted and then precipitated with two volumes of ethanol. The pellet was resuspended in 20  $\mu$ l TE (0.01 M Tris HCl, pH 7.5, 0.001 M EDTA).

#### B. Cloning in an Expression Vector

Lambda gt11 phage vector (Huynh) was obtained from Promega Biotec (Madison, WI). This cloning vector has a unique EcoRI cloning site 53 base pairs upstream from the beta-galactosidase translation termination codon. The genomic fragments from above, provided either directly from coding sequences (Example 5) or after amplification of cDNA (Example 4), were introduced into the EcoRI site by mixing 0.5-1.0  $\mu$ g EcoRI-cleaved gt11, 0.3-3  $\mu$ l of the above sized fragments, 0.5  $\mu$ l 10X ligation buffer (above), 0.5  $\mu$ l ligase (200 units), and distilled water to 5  $\mu$ l. The mixture was incubated overnight at 14°C, followed by in vitro packaging, according to standard methods (Maniatis, pp. 256-268).

The packaged phage were used to infect E. coli strain KM392, obtained from Dr. Kevin Moore, DNAX

(Palo Alto, CA). Alternatively, E. Coli strain Y1090, available from the American Type Culture Collection (ATCC #37197), could be used. The infected bacteria were plated and the resultant colonies were checked for loss of beta-galactosidase activity-(clear plaques) in the presence of X-gal using a standard X-gal substrate plaque assay method (Maniatis). About 50% of the phage plaques showed loss of beta-galactosidase enzyme activity (recombinants).

#### C. Screening for ET-NANB Recombinant Proteins

ET-NANB convalescent antiserum was obtained from patients infected during documented ET-NANB outbreaks in Mexico, Borneo, Pakistan, Somalia, and Burma. The sera were immunoreactive with VLPs in stool specimens from each of several other patients with ET-NANB hepatitis.

A lawn of E. coli KM392 cells infected with about 104 pfu of the phage stock from above was prepared on a 150 mm plate and incubated, inverted, for 5-8 hours at 37°C. The lawn was overlaid with a nitrocellulose sheet, causing transfer of expressed ETNANB recombinant protein from the plaques to the paper. The plate and filter were indexed for matching corresponding plate and filter positions.

The filter was washed twice in TBST buffer (10 mM Tris, pH 8.0, 150 mM NaCl, 0.05% Tween 20), blocked with AIB (TBST buffer with 1% gelatin), washed again in TBST, and incubated overnight after addition of antiserum (diluted to 1:50 in AIB, 12-15 ml/plate). The sheet was washed twice in TBST and then contacted with enzyme-labeled anti-human antibody to attach the labeled antibody at filter sites containing antigen recognized by the antiserum. After a final washing, the filter was developed in a substrate medium containing 33 µl NBT (50 mg/ml stock solution maintained at 4°C) mixed with 16 µl BCIP (50 mg/ml stock solution maintained at 4°C) in 5 ml of alkaline

phosphatase buffer (100 mM Tris, 9.5, 100 mM NaCl, 5 mM MgCl<sub>2</sub>). Purple color appeared at points of antigen production, as recognized by the antiserum.

5     D.   Screening Plating

          The areas of antigen production determined in the previous step were replated at about 100-200 pfu on an 82 mm plate. The above steps, beginning with a 5-8 hour incubation, through NBT-BCIP development,  
10     were repeated in order to plaque purify phage secreting an antigen capable of reacting with the ET-NANB antibody. The identified plaques were picked and eluted in phage buffer (Maniatis, p. 443).

15     E.   Epitope Identification

          A series of subclones derived from the original pTZKF1 (ET1.1) plasmid from Example 2 were isolated using the same techniques described above. Each of these five subclones were immunoreactive with  
20     a pool of anti-ET antisera noted in C. The subclones contained short sequences from the "reverse" sequence set forth previously. The beginning and ending points of the sequences in the subclones (relative to the full "reverse" sequence), are identified in the table  
25     below.

TABLE 1

	<u>Subclone</u>	<u>Position in "Reverse" Sequence</u>	
		<u>5'-end</u>	<u>3'-end</u>
5	Y1	522	643
	Y2	594	667
	Y3	508	665
	Y4	558	752
10	Y5	545	665

Since all of the gene sequences identified in the table must contain the coding sequence for the epitope, it is apparent that the coding sequence for the epitope falls in the region between nucleotide 594 (5'-end) and 643 (3'-end). Genetic sequences equivalent to and complementary to this relatively short sequence are therefore particularly preferred aspects of the present invention, as are peptides produced using this coding region.

A second series of clones identifying an altogether different epitope was isolated with only Mexican serum.

TABLE 2

	<u>Subclone</u>	<u>Position in "Forward" Sequence</u>	
		<u>5'end</u>	<u>3' end</u>
	ET 2-2	2	193
30	ET 8-3	2	135
	ET 9-1	2	109
	ET 13-1	2	101

The coding system for this epitope falls between nucleotide 2 (5'-end) and 101 (3'-end). Genetic sequences related to this short sequence are therefore also preferred, as are peptides produced using this coding region.

Two particularly preferred subclones for use in preparing polypeptides containing epitopes specific for HEV are the 406.3-2 and 406.4-2 clones whose sequences are set forth above. These sequences were isolated from an amplified cDNA library derived from a Mexican stool. Using the techniques described in this section, polypeptides expressed by these clones have been tested for immunoreactivity against a number of different human HEV-positive sera obtained from sources around the world. As shown in Table 3 below, 8 sera immunoreactive with the polypeptide expressed by the 406.4-2, and 6 sera immunoreacted with polypeptide expressed by the 406.3-2 clone.

For comparison, the Table also shows reactivity of the various human sera with the Y2 clone identified in Table 1 above. Only one of the sera reacted with the polypeptide expressed by this clone. No immunoreactivity was seen for normal expression products of the gtl1 vector.

Table 3  
Immunoreactivity of HEV Recombinant  
Proteins: Human Sera

Sera	Source	Stagel	406.3-2	406.4-2	Y2	λgt11
FVH-21	Burma	A	-	-	-	-
FVH-8	Burma	A	-	+	+	-
35 SOM-19	Somalia	A	+	+	-	-
SOM-20	Somalia	A	+	+	-	-
IM-35	Borneo	A	+	+	-	-
IM-36	Borneo	A	-	-	-	-
PAK-1	Pakistan	A	+	+	-	-
40 FFI-4	Mexico	A	+	+	-	-

FFI-125	Mexico	A	-	+	-	-
F 387 IC	Mexico	C	+	+	ND	-
Normal	U.S.A.	-	-	-	-	-

5 1A = acute; C = convalescent

While the invention has been described with reference to particular embodiments, methods, construction and use, it will be apparent to those skilled in the art that various changes and  
 10 modifications can be made without departing from the invention.